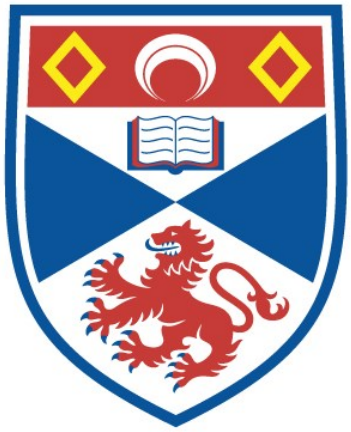
AN INVESTIGATION INTO MORITZ SCHLICK'S FOUNDATIONALIST EPISTEMOLOGY

Daniel James Healey

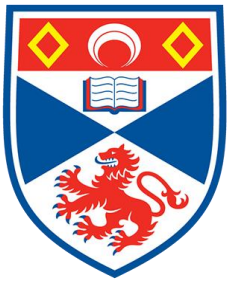
A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



2019

An Investigation into Moritz Schlick's Foundationalist Epistemology

Daniel James Healey



University of
St Andrews

This thesis is submitted in partial fulfilment for the degree of

Doctor of Philosophy (PhD)

at the University of St Andrews

March 2019

Abstract

Moritz Schlick is an influential figure in the history of philosophy, but his place in the narrative is often confined to having been the man who brought great thinkers together, rather than having been a great thinker himself. In this thesis I argue that Schlick's ideas deserve greater philosophical recognition, and to this end I focus on his work on the foundations of scientific enquiry. I trace Schlick's thought from *Allgemeine Erkenntnislehre* through the prism of Wittgenstein's *Tractatus Logico Philosophicus* and into his later work on the form and content of statements. I then look at the Vienna Circle's so-called "protocol sentences debate" and explain why Schlick felt the need to introduce his controversial account of *Konstatierungen*, his objective being to find epistemically-guaranteed foundations for our scientific beliefs. The problem with Schlick's account appears to be that any statement that is epistemically secure cannot be connected appropriately to our network of scientific beliefs, which itself is never immune to revision. I argue that Schlick may have been attempting to bridge this gap with the middle-Wittgensteinian notion of the criteria for the acceptance of a statement as separate from its truth conditions, but I argue that this approach leaves the link between *Konstatierungen* and science underexplained. Finally, I consider some of the advances made in philosophy since Schlick's death – Donald Davidson's arguments against the need for individually-infallible judgements to form the foundations of knowledge, and David Chalmers' scrutability framework which helps us explicate the connection needed between foundational statements and the system of science. I conclude that there is a viable position within the scrutability framework – "weak phenomenal structuralism" – that allows us to retain Schlick's emphasis on the role of experience in science and implies that science, as a whole, is well-founded, but individually-guaranteed *Konstatierungen* must stand wholly outside this system.

Candidate's declaration

I, Daniel James Healey, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 84,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree.

I was admitted as a research student at the University of St Andrews in September 2014.

I confirm that no funding was received for this work.

Date Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Daniel James Healey, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

No embargo on print copy.

Electronic copy

No embargo on electronic copy.

Date Signature of candidate

Date Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Daniel James Healey, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

Date Signature of candidate

Contents

Acknowledgements

This work would not have been possible without the endless patience of Peter Sullivan across countless hours of supervision. It is only through his ability to cut away nonsense and draw out arguments that what's left here has any semblance of coherence. As such, if there is anything of value in what follows, it can most properly be credited to him. Of course, Peter could only work with what I brought him, so I take full responsibility for the errors and contradictions that remain.

Parts of this work were presented at various seminars organised by the St Andrews and Stirling Graduate Programme. I would like to thank all the members of the philosophy postgraduate community for their helpful feedback from the Postgraduate "Friday" Seminar, and I would like to extend further gratitude to the members of the Arché philosophical research centre for valuable discussion there. I would especially like to thank the coordinators of the *Epistemology: Current Themes* research stream for torturing the meaning of "Current" to the point where I could present.

I would like to thank the University of St Andrews itself for providing office space, teaching opportunities, and access to otherwise-unobtainable academic resources. I couldn't ask for a better place in which to have spent these past few years of research.

Finally, I would like to thank Mr Browning, without whom I would have been an economist; Katrina, without whom I would have been lonely; and my parents, without whom I wouldn't have been at all.

Introduction

Friedrich Albert Moritz Schlick, 1882-1936, is one of those figures in the history of philosophy who don't get as much attention as they deserve. Paradoxically, he is also generally recognised as having been one of the most influential figures in the founding of the analytic tradition. Schlick's role in the common narrative is to have brought together a group of philosophers and scientists in Vienna whose collective work helped to kick off an ongoing, rigorous project of conceptual analysis, in opposition to what they saw as largely-meaningless metaphysical philosophy that was going on in Europe at the time. The work of this group would go on to influence almost every major figure in analytic philosophy, but despite this their own work is usually regarded as a failure, and Schlick in particular is better known for the effect he had on others than for any philosophy he advanced himself. In this thesis I will be reassessing Schlick's work with the aim of showing that he deserves to be more than merely historically influential – he deserves to be philosophically influential as well. To that end, there are two different projects running side-by-side through the following chapters. The first is an empirical historical project, which aims to faithfully reconstruct Schlick's position. The second is a philosophical project, which aims to either justify that position or make the smallest changes necessary to produce a justified position without losing sight of the motivations that produced it in the first place.

Of course, Schlick wrote about a broad range of topics during his life, and it would be a mistake to think that we could boil it all down to 200 pages without loss of substance. I will be focussing on the area in which Schlick had the most sustained interest and (in my view) the most interesting position– the intersection of language and epistemology. That which connects statements to beliefs is meaning, and it is the subject of Schlick's infamous claim: "the meaning of the proposition is the method of its verification" (1932b, [1979b, p.311]). This is sometimes presented as a dogmatic claim that the philosophers of the Vienna Circle put forward to rule out as meaningless any areas of discourse that they just didn't like, but it was in fact the end result of a rigorously-pursued line of inquiry, which we will go through in due

course. The exact nature of "verification", however, was to become the subject of a heated argument within the Vienna Circle, known as "the protocol sentences debate". The debate highlights the fact that the Circle was not a perfectly homogenous group as it is sometimes caricatured to be; in fact, there are very few positions which one can attribute the group as a whole as having held. There are elements of Schlick's philosophy that are unique to him and can be found in his writing from before he came to Vienna, and we will see that they play a crucial role in explaining his later position.

There are also unique elements of Schlick's philosophy that will *not* play a role in what follows, but which I wouldn't want to let pass completely without comment. Where many in Vienna thought that moral statements were meaningless (e.g. Neurath, 1931, [1983, p.56]; and one could make the case for including the Wittgenstein of *Tractatus Logico Philosophicus* in this category), or at best reducible to emotivism (e.g. Carnap, 1928, [1967, §152]; Ayer, 1946, [1952, pp.102-120]), Schlick wrote an entire book on the problems of ethics (1930, [1939]) in which he attempted to show that they could be meaningfully discussed as claims about happiness. He also wrote *Wisdom of Life* (1908) and "On the Meaning of Life" (1927) in which he argues that we can make purpose for ourselves by living with enthusiasm, finding joy in our day-to-day work. None of these works is particularly revolutionary, but they give us a glimpse of the person described by his colleagues as "modest, self-effacing, and kindly," with an "optimistic, roseate outlook" (Blumberg & Feigl, 1974, pp.xiii-xiv). Friedrich Waismann later wrote of Schlick:

> He possessed an openness to the rare and precious hours of this life, to those moments of illumination in which it seems to us as if a wall had broken down, and we mount to the skies and feel ourselves close to the living heart of the universe. When listening to a concert or gazing at a landscape, it could happen, in his own words, that the door to the infinite had suddenly opened for him. He saw in these moments the supreme meaning of life, and sought

for them as a lost and yet ever reoccurring trailmark of divinity.

- Waismann, 1938, [1979b, p.xv]

These poetic tendencies show up even in Schlick's more rigorous work on meaning, and occasionally obscure his point in ways that we will have to unpack as we go on. They do, however, make all of his essays a pleasure to read, which makes it even more of a shame that so few people continue to do so.

Feigl suggests that Schlick's optimistic outlook on life might have fuelled a disconnect with his students in later years, many of whom would have grown up during the Great Depression. Schlick himself was born to a well-off family in Berlin, 1882, where he remained throughout his education. He received his certificate of proficiency in 1900 from Luisenstädtisches Realgymnasium, where he had studied physics, mathematics, chemistry, and, almost as an afterthought, philosophy, becoming acquainted for the first time with the works of Descartes, Kant, Schopenhauer and Nietzsche. He would go on to study physics at the Friedrich Wilhelms University (nowadays known as the Humboldt University of Berlin) under future-Nobel-laureate Max Planck, writing his dissertation on the behaviour of light in non-homogenous mediums and graduating in 1904 (Schlick, 1912, [2006]).

Schlick had continued to pursue philosophy in his spare time while earning his doctorate, and he fully turned towards it very shortly after graduation. *Wisdom of Life* was ready for publication by the end of 1907, and in the winter semester of 1907/08 he moved to Zurich where he undertook a focussed study of psychology and philosophy, leading to his first philosophical papers. Their subject matter was primarily aesthetics, but Schlick's analytic style is evident in his rigorous analysis of just what the *question* of aesthetics amounts to. His foundationalist tendencies are visible from as early as page 3: "If we get quite clear on what it means to explain a fact, we see that it always amounts to displaying the latter as a special case of another fact… By continued explanation we are always bound to arrive eventually at facts which are incapable of further explanation," (1909, [1979a, p.3]).

After a failed attempt at the *Habilitation* in 1909, Schlick moved back to Germany, first to stay with family in Berlin and then to Rostock where he resumed

his philosophical work. Two years later he succeeded at his *Habilitation* there with his first major work in philosophy, "The Nature of Truth in Modern Logic" (1910, [1979a]). Here he was to put forward a theory of truth based in unique correspondence, the outlines of which would stay with him for the rest of his career (although we will note some development of the view by the time he went to Vienna). We also see here his earliest discussion of the importance of verification as the criterion by which judgements are determined to correspond with facts, although Schlick seems to put the cart before the horse by treating it as part of the analysis of what "truth" means rather than just the definition of "verification".

In 1915 Schlick published "The Philosophical Significance of the Principle of Relativity", and this was followed in 1917 by a short book, *Space and Time in Contemporary Physics: An Introduction to the Theory of Relativity and Gravitation*. Schlick's motivation with these was to demonstrate how Einstein's theories eliminate objectivity from space and time, as he wrote to Einstein in the extensive correspondence they shared at the time (letter from Schlick to Einstein, 04/02/1917, in Engler & Neuber, 2006, p.49; Oberdan 2013). The book proved popular and would go through four editions and an English translation in just the next five years. Einstein himself approved of the book and would contribute to Schlick's eventual appointment to the chair of *Naturphilosophie* at Vienna in 1922 – a post previously held by Ernst Mach.

In 1918, whilst he was still at Rostock, Schlick wrote the most substantial single publication that he would ever produce: *Allgemeine Erkenntnislehre*, later translated into English as *The General Theory of Knowledge*. It was incorporated into a collection of volumes on the natural sciences as the first in the series, dealing as it did with the epistemology that underlies the rest of scientific inquiry. This book, too, would undergo revision, and the second edition was published in 1925. That puts *Allgemeine Erkenntnislehre* earlier in the timeline than what could properly be called the Vienna Circle, and, although the *Tractatus Logico Philosophicus* was completed around the same time as the first edition, the consensus is that Schlick was not exposed to Wittgenstein's ideas until afterwards. There were embryotic

plans for a substantially-revised third edition, which would have been wonderfully useful for tracking the progress of Schlick's thought, but they were never realised (Feigl, 1938, [1979a, p.xxxv]).

The question of when exactly Schlick did come under Wittgenstein's influence is a matter of some debate in the literature, or at least a matter on which everyone asserts something different. Wittgenstein had published his *Tractatus* for the first time in 1921, and managed to get a version printed without the typographical errors in 1922. Schlick would have become aware of the *Tractatus* very shortly afterwards, as its thesis that the propositions of logic are senseless tautologies was discussed in a 1922 seminar by Hahn (Baker, 2003, p.xviii; Monk, 1991, p.213). The first time Schlick wrote to Wittgenstein, in which he calls himself "an admirer of [his] *Tractatus Logico-Philosophicus*", was in the summer of 1924 according to Monk (1991, p.241), and the two met in February 1927. (This was after a failed attempt to meet in April 1926, when Schlick arrived in Otterthal, where Wittgenstein had been working as a primary school teacher, only to find that Wittgenstein had just resigned and left town after his corporal punishment had caused a child to collapse (Monk, 1991, pp.232-233).) Feigl, one of Schlick's students and a member of the Vienna Circle, writes in the introduction to *Allgemeine Erkenntnislehre* (p.xvi) that the group undertook close readings of the *Tractatus* in both the 1924/25 and 1925/26 academic years, which is also the date given by Baker, 2003, p.xviii. Goldfarb, in contrast, says that the Vienna circle only read the TLP in the 1926/27 academic year (1996, p.219) and met him in the summer of 1927. And Oberdan (2013) dates it latest of all, saying that the first clear evidence Schlick had read the TLP in its entirety comes in a letter to Einstein, June 1927. Schlick's first academic reference to the *Tractatus* occurs in 1926 and refers specifically to the 1922 publication. As a historical question, this is interesting, and it might be possible to tease some Wittgenstein scholarship out of it. (For example, are there positions which we are unsure of attributing to Wittgenstein that show up in Schlick's 1926 work? If he's read the *Tractatus* at this stage and Wittgenstein doesn't accuse him of failing to understand the work, then that's relevant. If he hasn't read it yet, it's of no consequence.) It's certainly relevant for an evaluation of

Schlick's work in the 1924-1927 period. But I will be sidestepping the issue by not discussing the works of this period in serious depth (though this is certainly not to say that I am ignoring them altogether). My focal points will be *Allgemeine Erkenntnislehre* as Schlick's pre-Wittgenstein philosophy (it received only relatively minor revisions between the 1918 and 1925 editions, so I don't consider the small overlap problematic), and "Form and Content" (1932b, [1979b]) as a definitively post-Wittgenstein position which is further developed in the years that followed.

Around the same time as the Vienna Circle was undertaking its readings of Wittgenstein, Carnap was writing his *Konstitutionstheorie*. This would become the most famous work produced by the Vienna Circle, in spite of being written almost entirely before they formed in earnest: *Der Logische Aufbau der Welt*, or *The Logical Construction of the World*. Schlick was in correspondence with Carnap towards the end of its production, as Carnap was preparing to submit part of it for his *Habilitation* at Vienna (Uebel, 2007, p.34), and references the forthcoming work as early as 1926. It was shortly after the *Aufbau*'s publication in 1928 that the group that had been calling themselves the Schlick Circle rebranded to The Vienna Circle, and it was the year after this that the first issue of *Erkenntnis* was produced, detailing the group's research for the year. The opening essay of the journal was Schlick's "The Turning Point in Philosophy" (1930a, [1979b]) in which he laid out the Circle's view that there was no such thing as a "philosophical truth", and that the subject matter of philosophy was restricted to the clarification of concepts – an insight which he attributes to Wittgenstein and which he thinks will precipitate "an altogether final change in philosophy" (p.155). It would be just a couple of years after this optimistic start, however, that in the pages of *Erkenntnis* the Vienna Circle would begin to tear itself apart.

The construction of the world in the *Aufbau* reduced physical and psychological concepts to phenomenal concepts with a relation of recollected similarity, but never explicitly talked about how statements about immediate observations would be formed. A debate arose in the Circle over whether these

statements would refer to objects in the world ("there is a red cube on the table") or phenomenal concepts ("a red patch is next to a brown patch"), and what the epistemic status of these statements would be. Schlick was adamant that the system of science should be grounded in epistemically-secure statements, while Neurath in particular argued that it was a condition of something being a meaningful sentence that it *not* be immune to revision. The protocol sentences debate ultimately exposed cracks in the logical positivist programme which could not be easily filled, and with the approach of WWII the members of the Vienna Circle were scattered. Many went to the US and UK, where their influence would be unquestionable in the development of anglophone philosophy, but Schlick did not live to see this.

The protocol sentences debate was ended abruptly when, on 22nd June 1936, former student Johann Nelböck shot Schlick four times with a pistol as he ascended the steps of the university. Schlick died at the scene. Nelböck had previously been admitted to a psychological institution twice with schizoid personality disorder, and Schlick had, for a time, had to employ a bodyguard in response to his death threats, but he was found fully compos mentis at trial and sentenced to ten years in prison. He was released two years later after Austria's Anschluss with Nazi Germany, arguing for a pardon on the grounds that he had helped prevent the spread of Jewish philosophy. (Schlick was not actually Jewish himself, but his refusal to denounce Jewish friends such as Waismann had angered the Austrian fascists and may have made them generally receptive to this line of reasoning.) At his trial, Nelböck had claimed that he did what he did because Schlick's philosophy had undermined his moral restraint, but privately he seems to have been motivated by the delusion that Schlick was responsible for failures in his romantic life. The Austrian senior prosecutor judged that Nelböck had acted out of private motivations rather than political ones, but nevertheless released him on probation.

The Structure of this Thesis

The first chapter of this thesis is focussed on the pre-Wittgenstein *General Theory of Knowledge*, with a particular focus on the way definitions are cashed out. Schlick realised that the meaning of scientific terms couldn't be tied to the content of our experiences without being made too imprecise for scientific use. The precise

meaning of a term can only be properly given in a definition, which is to say that meaning is bound up in the relations between terms, rather than a relation between a term and a thing. Here we get the first glimpse of the structuralist position which is going to emerge. Schlick likens the definitions of terms by their relations to one another to that of mathematical concepts. Concepts like *successor* are defined by their role in the Peano axioms, rather than by any intuitive grasp of what it is for one thing to follow another. Likewise, Schlick thinks, "time" is defined by its role in the theory of relativity rather than by any intuitive grasp on the phenomenal content of time.

In the second chapter I introduce a reading of Wittgenstein's *Tractatus* and look at the ways in which Schlick's position can be seen to shift under its influence. A lot of the views put forwards in the *Tractatus* are already present in Schlick's earlier philosophy, although Schlick thought that his own work was amateurish in comparison to it. The most significant advance which we'll pick up on at this point is Schlick's adoption of the Picture Theory of language. Schlick already espoused a correspondence theory of truth, meaning that a sentence is true iff there is a fact in the world which corresponds to it, but the manner in which this correspondence relation was set up changed after the *Tractatus*. Previously, Schlick had said that "a judgement pictures the nature of what is judged as little as a musical note pictures a tone," (1925, [1974, p.61]), and the correspondence was taken to be a feature of the concepts in the statement being instantiated by the same single object (whereas in a false statement, like "all dogs are black", say, we find that the statement does not correspond to a single fact because "dog" and "black" pick out distinct sets). Afterwards, Schlick endorses the view that the elements of a sentence are arranged in a way that matches the structure of the objects in the fact. Schlick's structuralism becomes a stronger component of his account of language because, more than merely guaranteeing the precision of science, now it helps to explain truth. We also find the last vestiges of phenomenal content disappearing from the account at this time, as everyday words like "blue" come to be defined by their structural properties rather than by whatever experience we associate with the word.

The third chapter is the most significant in terms of constructing a defensible account of Schlick's position. Here we look at the views laid out in "Form and Content" – probably the most well-developed version of Schlick's structuralism, albeit an account plagued with poetry. In particular, he is sometimes accused of attempting to express the inexpressible in this essay. Schlick, at the time, was committed to Wittgenstein's view that it was impossible to get outside of language to talk about its rules. Truths about language could not be *said*, they could only be *shown*. This meant that the only way for Schlick to get his view across was to use terms that his own view said were nonsense. For Schlick, "Content" refers to that which is not communicated in the structure of a statement, and can be compared to what philosophers today would call "qualia". *That there is something which cannot be said*, seems like the kind of metaphysical claim you might find in Bergson or Husserl, but I will argue that Schlick was not trying to argue for inexpressible facts. On the contrary, the reason that Content cannot be expressed is that it is not a fact, and therefore not the kind of thing which you would ever want to say. That is, Form and Content is not a *distinction* in the sense that, say, Analytic and Synthetic is a distinction amongst sentences – Form and Content are totally different categories of things, like Analytic and Mammal. Trying to express content, I will argue, is like trying to express a rock. You can write down and say everything there is to say *about* the rock, but the rock itself won't go anywhere. Nevertheless, "there is a rock" is not a metaphysical claim.

Chapter 4 is an account of what was happening elsewhere in the Vienna Circle around the same time as "Form and Content"; namely, the protocol sentences debate. We start with an account of Carnap's *Aufbau* (in fact, three accounts) and look at the role that protocol sentences were thereby set to play in the Circle's philosophy. Then we look at the objection from Neurath, already alluded to, which threatened, in Schlick's eyes, to sever the whole project's connection to empiricism. Neurath's objection was, roughly, that if protocol sentences are meaningful then they are open to disconfirmation. Nowhere in the system of science can we compare statements with facts directly, because all we have to go on is what we believe to be true. We can't get outside our own network of beliefs in order to see whether one of

the beliefs is *really* true. On one account, which appears to have been Schlick's interpretation, Neurath was arguing that truth is determined not by correspondence but by coherence of the system. We will see, however, that Neurath was really rejecting the notion of truth altogether as a metaphysical notion with no place in science.

Chapter 5 considers Schlick's response to Neurath and brings us back to his epistemology. Schlick attempts to ground the system of science epistemically in first-person present-tense statements of private experience – *Konstatierungen* or "affirmations" – which constitute the moment of verification for a theory and provide the impetus for forming new theories to predict future experiences. This is the most controversial part of Schlick's philosophy, directly in contention with Neurath's claim that any sentence which is appropriately connected to the system of science must be vulnerable to disconfirmation. There are a couple of different interpretations in play to explain what Schlick was trying to do. We will look at the differences between the accounts of Thomas Oberdan and Thomas Uebel. On Oberdan's account, we should ignore Schlick's emphasis on certainty and construe affirmations as sentences in a phenomenal language which can nevertheless be translated into the language of science. On Uebel's, affirmations stand outside the language of science as infallible, connected to scientific enquiry not by an implication relation but by a psychological motivation relation; Uebel argues that Schlick was, under the influence of Wittgenstein, beginning to see language as a living entity rather than a rigid calculus, and so there were aspects of it which would not connect directly to the strict formal system of Carnap. I'll be arguing that Uebel's account meshes best with what Schlick was saying in the years following the introduction of affirmations, and with the contemporaneous work of Wittgenstein with which we know Schlick to have been familiar, and I will argue that developing Wittgenstein's notion of the criteria for justified assertions as separate from their truth conditions seems to elucidate how Schlick thought his infallible affirmations might relate to the fallible system of science. I will nevertheless argue that the criterial relation is underexplained and doesn't give us good reason to think that affirmations can serve to justify scientific beliefs in the way Schlick had hoped.

Indeed, I will say that nothing could achieve Schlick's stated objectives, for reasons much like those raised by Neurath.

In chapter 6 we turn to look at advancements made in more recent years by philosophers who were well-informed of the Circle's internal struggle. The two figures who form the focus of this chapter are Donald Davidson and David Chalmers. Davidson provides an account of meaning that does not depend on the subject's experiences at all, they being inaccessible to an interpreter; he goes on to argue that we shouldn't expect there to be individual firmly-grounded statements to epistemologically found the system of science, but also that we can nevertheless expect the system to correspond in the most part to the state of the world, and therefore to be true. Chalmers' more-recent work I take to be able to respond to some of Davidson's arguments and help us get back to something more like Schlick would have wanted. Chalmers provides us with a new method of philosophical construction, weaker than that of Carnap but still strong enough to give us many of the conclusions we want, and resistant to the objections which plagued the *Aufbau*. Chalmers construction is a metaphysical one rather than an epistemic one, so works towards different objectives than Schlick, but it does give us a clear relationship between propositions such that if we can find good grounds to accept the propositions in a particular class then we will have good grounds to accept the system of science as a whole. Chalmers gives us a response to global scepticism which suggests that, assuming we can justify an initial move to structuralism, sceptical hypotheses will never get a purchase on us. That allows me to conclude, on the basis of arguments made in chapter 3 (and a weakening of that position in response to some other issues that Chalmers brings up), that Schlick's attempt to establish firm foundations for the system of science *as a whole* in empirical sense data is not a failure. Schlick's structuralist account gives us a base from which the entirety of science can be derived, which no individual can know in its entirety but also about which no individual can have entirely false beliefs, and which appears to derive its justification from our immediate experiences.

Our conclusion will be that Schlick gives us a well-motivated structuralist account of language, and once that structuralist account is in place there are epistemic conclusions that follow naturally. It does seem as though there can be no statements within the system that have a stronger epistemic status than the system as a whole, but the system as a whole nevertheless seems to be relatively secure, and the propositions that make up the system appear to hold their justification in virtue of their link to immediate experience. In achieving this, we weaken relations between propositions to a point where the *meaning* of statements cannot be formed from statements in an epistemically foundational class, and so the verification principle as stated does not hold. Nevertheless, we claim that there are no truths which are isolated from claims about experience, and the spirit of the principle survives.

Abbreviations

The following works were cited often enough in this thesis to warrant abbreviation. The abbreviations will be introduced again in the text, but I insert this here for easy reference.

AE[1] - Schlick, M., 1925, [1974], *Allgemeine Erkenntnislehre*, translated as *General Theory of Knowledge*, trans. Blumberg, A.E., ed. Blumberg, A.E., & Feigl, H., Open Court, LaSalle, IL

F&C - Schlick, M., 1932b, [1979b], "Form and Content. An Introduction to Philosophical Thinking", *Moritz Schlick, Philosophical Papers, Volume II, [1925-1936]*, ed. Mulder, H., & van de Velde-Schlick, B.F.B., D. Reidel Publishing Company, London, pp.285-369

UFE - Schlick, M., 1934, [1979b],"Über das Fundament der Erkenntnis", *Erkenntnis*, Vol.4, pp.79-99 reprinted as "On the Foundation of Knowledge", *Moritz Schlick, Philosophical Papers, Volume II, [1925-1936]*, ed. Mulder, H., & van de Velde-Schlick, B.F.B., trans. Heath, P., D. Reidel Publishing Company, London, pp.370-387

TLP[2] - Wittgenstein, L., 1922, *Tractatus Logico-Philosophicus*, trans. Ogden, C.K., Kegan Paul, Trench, Trubner & Co., Ltd, London

---

[1] References to AE will always be to the English translation of the second edition unless clearly stated otherwise.

[2] References to the TLP will be to individual statements using Wittgenstein's numbering, e.g. (TLP: 2.0321) refers to the statement labelled 2.0321, which is on p.30 of my copy. Similarly, references to the *Philosophical Investigations*, *Philosophical Grammar*, or *Philosophical Remarks* will be to particular remarks rather than pages.

1. Definitions in Early Schlick

1.0 Introduction

Schlick began his academic career as a physicist, and his philosophical work sprang from a desire to see the sciences rest on a firm foundation. Indeed, his first major work was a philosophical treatment of relativity in reaction to those philosophers, in particular Kantians, who rejected it as a treatment of space and time because it merged these two intuitively-distinct notions. As science is the pursuit of knowledge it appeared to be necessary to establish exactly what knowledge was, and this was the goal of Schlick's 1918 work *Allgemeine Erkenntnislehre* (*General Theory of Knowledge*; 2nd edition 1925; henceforth "AE"; citations will refer to the pagination of the English translation of the 2nd edition unless otherwise noted). As we will see in this chapter, Schlick's considerations on the nature of knowledge led him to ideas about definitions which would be fundamental in the development of his later philosophy.

In the first section I will explain how considerations of knowledge lead us to considerations about definitions, and we will see why Schlick found the traditional empiricist conception of definition (in terms of experience, or "intuition") inadequate. In the second section I will look at how Schlick hoped to use implicit definitions, a notion borrowed from Hilbert's mathematics, to solve the problems which intuitive definitions raised. I will go on to examine the problem which Carnap raised for implicit definitions, and we will see in the next couple of chapters how Schlick's view changed under the influence of Carnap and Wittgenstein.

1.1 Concrete Definitions

At the time of AE, Schlick was not concerned with the possibility of language so much as the possibility of knowledge. The philosophy of AE is always in terms of

judgements and concepts rather than propositions and words. This leads to some very different arguments than we find in Schlick's later work, despite the conclusions being strikingly similar. Schlick uses the word “knowledge” in a broad way, covering all the judgements science might make about the world. The important feature is not certainty, or even justification, but interconnectedness. Knowledge is about reducing one thing to another – for instance, you know what Fido is if you’re able to associate Fido correctly with the previously-known concept *dog*. In this way, most terms are defined explicitly by their relation to other terms. But the interesting question is how this interconnection terminates. Importantly, Schlick rejects the idea that we might have to make do with “only approximate or probable knowledge” (AE, p.30), and so he is seeking absolute precision and certainty for the foundations of the network, although of course higher-level scientific propositions will remain hypotheses. In contrast to the language of science, Schlick thinks that everyday language tolerates imprecision and vagueness. This separation of everyday language from the language of science will prove problematic in section 1.2.5, where the difficulties of linking Schlick’s model of scientific terms to experience will become apparent.

In this section I will run through the line of reasoning Schlick followed in AE in his first attempt at establishing the precision and certainty of the foundations of science. In particular we will be looking at the importance of the notions of Concept and Intuition. These were the forerunners of Schlick's later ideas of Form and Content, which we will cover in more detail in chapter 3. It may seem at first that Intuition, in some sense, underlies Concept, and various continental philosophers in the early 20th century - notably Bergson and Husserl - had argued that intuitions are the foundation of knowledge. As we will see, Schlick was to argue that we could do away with them entirely. Crucially, Schlick would argue that knowledge always has two components – the object which is the subject of the knowledge, and the concept that the object is known to fall under.

1.1.1 Concepts

Schlick started from the view that to know (*Erkennen*) an object is to be able to recognise it as something already familiar. Suppose we are presented with an animal, which we'll call "Fido", that we want to study and learn things about. When we know something about Fido, we know *Fido is a dog*, or *Fido is brown* or *four-legged* etc.. In Schlick's words it is "to equate *what* is known with *that as which* it is known," (AE: p.15). This act of equating, however, creates a puzzle. "Equating presupposes comparing. In acquiring knowledge, what do we compare with what?" (AE: p.15).

We obviously don't compare the object itself to the property. Neither Fido nor the property *x is a dog* are present in my head. In everyday life, Schlick says, it seems that what we compare are images. We have an image of Fido from having looked at him. We have an image of what a dog looks like from our memories of other dogs. We compare the two and if they match then we form the judgement that Fido is a dog. However, as Schlick points out, this just creates further problems. For one thing, an image from memory is generally fuzzy. The memory image I have of my house is not specific as to how many bricks it is built out of, so how is it that I am able to look at a building with a determinate number of bricks and be certain that it is the same as the house in my memory? The case of recognising Fido as a dog is even more difficult, because there are lots of different ways an object can look whilst still looking like a dog. I can recognise Fido as a dog if he looks like a poodle and I can recognise him as a dog if he looks like a corgi, so does my mental image of *x is a dog* resemble a poodle or a corgi? Non-visual images are even more difficult, because humans aren't good at recalling other kinds of experience. The smell of a rose is different to that of a daffodil, but you'd be hard pressed to recall the precise difference without actually being presented with the two.

We might think that we can avoid the fuzziness problem by suggesting that the images involved in everyday recognition will be more or less archetypical, and seeing whether or not two images match will be a question of similarity rather than identity (AE: p.18). When we conjure up an image for the property *x is a dog* we

primarily have a specific image in mind, be it a poodle or a corgi or whatever, but this image comes with the secondary thought *this is not the only kind of dog*. Our experience of dogs tells us the various ways an animal can vary from the archetype whilst still being a member of the species, so we know that an animal having white fur is no reason to discount it from being a dog even though the archetype has brown fur, while an animal with eight legs is quite unlikely to be a dog. Schlick maintains that this process is more or less how acquaintance goes in everyday life. He does, however, acknowledge that the process is not infallible. Clearly there are cases in which we are mistaken about the nature of the thing in front of us. There is a wide variety of dogs in the world and some of them will be very different from our archetype indeed. If you showed a poodle and a fox to a person whose only prior acquaintance with dogs had been with corgis, then you would be hard pressed to convince them that the poodle was a dog and the fox wasn't. For this reason, Schlick recognises the need for a more rigorous set of criteria for properties like *x is a dog*. Clearly if we admit that people going by similarity-to-archetypes can be wrong about whether or not something is a dog then there must be something besides similarity-to-archetypes by which dogness is measured. This more-rigorous something is the concept.

> A concept is to be distinguished from an intuitive image above all by the fact that it is completely determined and has nothing uncertain about it… Thus a concept is not an image. It is not a real mental structure of any sort. Indeed, it is not real at all, but imaginary… We operate with concepts as if they were images with exactly delineated properties that can always be recognised with absolute certainty. These properties are the *characteristics* or *features* (*Merkmale*) of the concept, and are laid down by means of specific stipulations which in their totality constitute the definition of the concept… Thus it is through definitions that we seek to obtain what we never find in the world of images but must have for scientific knowledge: absolute constancy and determinateness.
> - AE: p.20, original emphasis

The first point to notice is that concepts are to be treated similarly to images. For Schlick, the scientific act of recognising, for example, Fido as being a dog still involves comparing a sense-data image of Fido with *something*. Whatever the concept of *x is a dog* is, it still has to act like the archetypal image of a dog that we have in everyday recognition. But nothing in the mind can do this job with the required clarity, hence why concepts are not mental structures. In fact, Schlick does not appear to think that there is anything which can properly do the job - in the passage above he calls them imaginary; later he puts them in the category of "arbitrary fictions" (AE: p.21).[3]

A concept does not in fact exist, but it has characteristics in the sense that there are real properties which an object must possess in order to fall under it, and these characteristics form the underlying structure of knowledge in science. Rather than rely on the vague notion of sufficient similarity to a blurry image, scientific knowledge is grounded in precise definition. Fido is a dog not simply because he resembles other dogs - he is a dog because his properties properly correspond with the characteristics of the concept of *x is a dog*, say, that he is a four-legged, mammalian, etc., or that he has a certain kind of DNA.[4] We recognise Fido as a dog in everyday life by his resemblance to our archetype, but experts decide whether or not our first guess was right by checking whether Fido's properties line up with the characteristics of the concept of *x is a dog*.

---

[3]Schlick mentions logic and mathematics as being entire sciences "whose subject matter consists exclusively of concepts and their relationships" (AE: p.24). Schlick accepts that we may want to talk about things like numbers and triangles as though they exist, perhaps attributing to them an "ideal" existence as opposed to a physical one, and makes it clear that he has no problem with people adopting this terminology if they wish, so long as they do not thereby confuse themselves into questioning the relationship between two realms of existence or other such pseudo-problems. His claim that concepts do not exist is restricted to denying their physical existence as structures in the brain or otherwise as mental entities.

[4] In practice it's more difficult to say exactly what features Dog has, but explicitly drawing them out is beyond the scope of this project.

Comparing the object to the concept is how we know that the thing before us has a particular property. Knowing consists in recognising that one thing is an instance of another, or, equivalently, being able to designate the object by its correct name(s). At its most basic, to know Fido is to know that *this* is correctly called "Fido", where being correctly called "Fido" consists in exemplifying the characteristics of the concept of *Fido*. The use of a concept is a conceptual function. Concepts themselves do not exist but conceptual functions do - they are acts which can be carried out through mental images or by written signs. In the language of science every word (aside from the purely grammatical ones like "than") is supposed to designate a concept (AE: p.21). In everyday language a mental image with fuzzy boundaries is sufficient for most purposes. Concepts in AE are supplemented by the early analogue of Content – Intuition:

1.1.2 Intuition

Concepts are sufficiently precise that they might be used to accurately represent the world in a way that perceptual images cannot, but Schlick points out that the vagueness inherent in perceptual images is still going to cause us a problem when it comes to knowledge. That's because it seems as though we're going to have to make use of perceptual images in the activity of recognition.

Suppose again that we are interested in examining Fido, and in particular at this moment we are interested in the colour of his fur. When we say that we have only a perceptual image of Fido to go on, this is not in the same way as if we were to try to identify him from a photograph. We have the dog in front of us and we can conduct a thorough inspection of it. We can make sure it has not had its coat dyed. We can check it under various different lights and against various different backgrounds to get as unbiased a view as possible of the colour of its fur. All the empirical investigations we can possibly undertake will contribute to our perceptual image. The problem here is therefore not that our perceptions might be inaccurate, or that we might be deceived. The epistemic problem that Schlick is aiming at here is not a shortage of evidence, but something more fundamental. It's that perceptual images can *never* attain the kind of precision that scientific inquiry demands. We

might idealise our investigation by saying that we examine Fido under “ideal lighting conditions” or with a colour comparison chart in hand, but we know enough about colour-based Sorites paradoxes to know that we will never get as clear an image of Fido’s colour as our concepts demand.

Now suppose we are instead interested in Fido's DNA, we can hardly investigate that from raw perceptual imagery, but our investigation may be carried out with the aid of instruments. We can take samples of Fido's blood, run them through a machine, and read the results off the monitor. In this way we drastically reduce the number of ways in which we can go wrong. "In the final analysis, however, sensory observation, such as the reading of a scale, always involves recognition of a perceptual image, and the latter, as we have made clear, is ever subject to an essential uncertainty," (AE, p.27) - although we can establish whether or not Fido is a dog with sufficient certainty for practical biological purposes, this is not enough for a theory of precise scientific knowledge. Even if our concept of *dog* is equivalent to *produces a DNA analysis that looks like this*, the “looks like this” part introduces intuitive imagery that undermines the precision. Ultimately recognising an object as falling under a particular concept appears to rest, at some point, on recognising sense data.

Schlick points out that, given the reliance of concepts on definitions, we can arrive at this problem by considering concepts in the abstract. Concepts are defined in words, and all words in the language of science designate concepts so they must themselves be defined. This process will lead to an infinite regress unless we can introduce terms somewhere without verbal definition. The obvious candidates are words like "blue" or "pleasure", which "we cannot learn … by definition, but only by intuiting something blue or experiencing pleasure. … [A]n eventual return to what is immediately given, to intuition and experience, is unavoidable," (AE: p.29). But we have seen that immediate experience is fuzzy, and not something with which we can ground absolutely precise concepts. Other philosophers in the early 20th century such as Henri Bergson and Edmund Husserl had argued that Intuition - immediate experience - is the source of knowledge, but Schlick firmly disputes this.

Bergson's (1903, [translation undated]) argument runs something like this, insofar as it is possible to read a coherent argument from Bergson: Saying something about the world means translating your experience into symbols (e.g. sounds or written words). This translation is always imperfect. If we wanted to perfectly give the coordinates of an object's location, for example, we would need to use an infinite number of digits because space is not discrete, but this is obviously impossible. Therefore, a description of the world will never be as good as experiencing it. Experience - that is, Intuition - is therefore the only way we can arrive at accurate judgements about the world. Accuracy is needed for truth which is needed for knowledge, so knowledge is grounded in Intuition.

Schlick's response is that what we get from Intuition is not knowledge in the sense in which we're interested, because the mere experience of something does not involve attributing a predicate of a subject. It doesn't allow us to form a judgement of any kind, let alone a true and certain one, because it doesn't place a concept into a network of relations with other concepts – it doesn't *say* anything. He does say that he has no objection to people using the word "knowledge" like this if they like and, if so, we will have to distinguish between "intuitive knowledge" and "conceptual knowledge" and only talk about the latter. If we're looking for something which can be of use to science, we will need to look beyond Intuition.

The idea that knowledge (*Erkennen*) always involves two things ("*something* that is known and *that as which* it is known" (AE, p.82, original emphasis)), while acquaintance (*Kennen*) involves just one, was one which would stay with Schlick throughout his later philosophy. In 1932's "Form and Content" (1932b, [1979b], henceforth "F&C") he says:

> Bertrand Russell distinguishes between 'knowledge by acquaintance' and 'knowledge by description', but why should the first be called knowledge at all? The word 'acquaintance' alone seems to me to be sufficient, and then we can emphasize the distinction between *acquaintance* and *knowledge*. There is no

similarity of meaning between the two.

-F&C, p.318, original emphasis

This means that the intuitive experience we associate with a scientific concept is not necessary for that concept's definition, and it is not even part of what we know about objects when we know that they fall under that concept.[5] To define these concepts without recourse to experience, Schlick introduced implicit definitions.

1.2 Implicit Definitions

We have seen that the only way that concepts can be sufficiently precise for use in science is when they are defined, as opposed to merely being associated with intuitive images. "To define a concept is to specify its characteristics. But these latter, if they are to be precisely determined, must in turn be defined; that is, they must be resolved into further characteristics," (AE, pp.28-29). However, if these definitions are not to be circular it seems like they have to stop somewhere. The traditional empiricist answer is to reduce complex concepts to simple ones which can be defined ostensively, but we have seen that this was not good enough for Schlick. Ostensive definition reintroduces the imprecision that this process was meant to get rid of in the first place. Instead, he looked to the work of David Hilbert, whose 1899 work on the foundations of geometry seemed to provide the solution to Schlick's problem.

[5] For many of the precise scientific terms Schlick is interested in, this won't be a surprise. Most of us don't have any intuitive preconceptions about the meaning of the term "electron", for example. But another term that will be defined without recourse to experience is "time" – Schlick was keen to make it clear to people that Einstein's theories could not be rejected merely because our intuitive experience of time involves an objective present and uniform passage regardless of speed. This was the subject of Schlick's most famous early work, which predated AE: *Space and Time in Contemporary Physics* (1920; the first edition was published in 1917).

In this section I will explain Hilbert's use of axioms in the definition of geometric terms and how Schlick sought to expand this approach to definitions to other areas of language. Implicit definitions retained an important place in Schlick's philosophy until 1926, after which they disappear abruptly. In section 1.3, we will examine the objection from Carnap, whose 1927 paper "Proper and Improper Concepts" attacked the legitimacy of using implicitly defined concepts in an empirical theory. This will lay the groundwork for our forthcoming discussion of Wittgenstein, whom Schlick first met in person in February 1927, and whose philosophy helps to fill in some of the gaps in the story of implicit definitions.

1.2.1 Hilbert & Frege

In 1899 Hilbert published *Grundlagen der Geometrie* (*Foundations of Geometry*) in which he laid out a set of axioms for Euclidean geometry and demonstrated their mutual consistency. His proof consisted, in part, in translating the geometric axioms into statements about the domain of real numbers, the consistency of which is assumed (the consistency proof is relative to the assumed consistency of arithmetic). Thus, points are treated as coordinates (x, y), lines are defined by functions such as $y = mx + c$, and so on. Frege considered this move illegitimate, and wrote to Hilbert to express his concerns.

Frege's main objection was that terms like "point" and "straight line" have meanings which extend beyond their formal relations. "Point" and "pair of numbers" are fundamentally different concepts, and it would be misleading to talk of them as being somehow interchangeable. It might be true that some statements about geometry can be interpreted using real numbers, but that is not sufficient to make them statements *about* real numbers and we shouldn't necessarily think that reasoning using real numbers will lead us to new geometric truths.

Patricia Blanchette illustrates Frege's point as follows:

> Consider the set of sentences {Jones had a nightmare, Jones didn't have a dream}, or equivalently its first-order rendition, {*Nj*, ~*Dj*}. The set is clearly consistent in the sense used by Hilbert in [*Foundations of Geometry*]; it is a straightforward matter to

> provide interpretations of “Jones,” “*x* had a nightmare” and “*x* had a dream” (or of “*j*”, “*N*”, and “*D*”) such that the sentences, so interpreted, express truths. (Consider, for example, an interpretation on which “*j*” is assigned the number 7, “*N*” the set of prime numbers, and “*D*” the set of numbers greater than 12.) But from the Fregean point of view, the thoughts expressed are inconsistent, since part of what it is to have a nightmare is to have a dream. The inconsistency from Frege's point of view can be demonstrated by providing an analysis of the thoughts expressed, and noting that the results of this analysis yield the set {Jones had a disturbing dream, Jones didn't have a dream}.
>
> - Blanchette, 2012, §4

That is, if we don't pay attention to the concepts which the terms represent, the sentence "Jones had a nightmare" is perfectly consistent with the sentence "Jones didn't have a dream". "Nightmare" and "dream" are different words - there's no formal reason why you couldn't have something called "nightmare" without having something called "dream". But if we understand "nightmare" as being the term for disturbing dreams then it's clear that the two sentences are not consistent - one cannot have a nightmare without having a dream. Frege was concerned that the same kind of problem might arise in Hilbert's geometry. It's possible for some feature of points and lines to go unrepresented in the real-number interpretation of the geometric axioms, and it's possible for such a feature to falsify the conclusion of a line of reasoning. Frege didn't have any particular feature in mind and he found no specific problem with Hilbert's proofs; his objection was merely that they were based on an unjustified move.

Hilbert disputed Frege's claim that geometric terms have any meaning beyond that which is given in the axioms.

> If one is looking for other definitions of a 'point', e.g., through a paraphrase in terms of extensionless, etc., then I must indeed

> oppose such attempts in the most decisive way.
> - Hilbert, letter to Frege, 29/12/1899, in Frege, 1980 p.39

> In my opinion, a concept can be fixed logically only by its relations to other concepts. These relations, formulated in statements, I call axioms, thus arriving at the view that axioms (perhaps together with propositions assigning names to concepts) are the definitions of the concepts,
> - Hilbert, letter to Frege, 22/09/1900, in Frege, 1980, p.51

Hilbert's view was that "point", "line", and so on, are defined entirely by the axioms in which they occur. A point is anything which, in conjunction with certain rules, satisfies the axioms, so a pair of numbers (x, y) can be a point so long as you also define a line (e.g. Hilbert uses $(u : v : w)$) and have a rule by which the point lies on the line (e.g. "the point (x, y) lies on the line $(u : v : w)$ when $ux + vy + w = 0$"), a rule by which one point is between two others, and so on.

1.2.2 Support for Structuralism

This account of the meaning of mathematical terms is well-motivated. Benacerraf (1965, [1983]) argues that if we think of numbers as sets, as Frege did, then we are presented with a number of equally good options for *which* set any number might be. The number 3, for example, can be constructed as {{∅}, {∅, {∅}}, {∅, {∅, {∅}}}} or as {{{∅}}}, and in either case it is easy to interpret it so that all of the axioms of Peano arithmetic hold. Frege took the number 3 to be (*very* roughly) the set of all sets with cardinality 3, but Benacerraf argues that "[a]lthough an appealing notion, there seems little to recommend it over … [{{∅}, {∅, {∅}}, {∅, {∅, {∅}}}}]" (1965, [1983, p.281]). The question of which set, if any, is really the extension of "3" need not concern us here. Benacerraf's summary of the structuralist view of mathematics provides us with a good idea of what we will need in order to extend the account, as Schlick did, to language as a whole:

> If numbers are sets, then they must be particular sets, for each set is some particular set. But if the number 3 is really one set rather than another, it must be possible to give some cogent reason for

> thinking so; for the position that this is an unknowable truth is hardly tenable. But there seems to be little to choose among the accounts ... *for the accounts differ at places where there is no connection whatever between features of the accounts and our uses of the words in question*.
> - Benacerraf, 1965, [1983, pp.284-285], original emphasis

> [N]umbers are not objects at all, because in giving the properties (that is, necessary and sufficient) of numbers you merely characterise an abstract structure - and the distinction lies in the fact that *the 'elements' of the structure have no properties other than those relating them to other 'elements' of the same structure.*
> - Benacerraf, 1965, [1983, p.291], emphasis added

That is, the way in which we use things like numbers and points in arithmetic and science does not, in general, depend on the specifics about how we conceive of those things. It doesn't matter what set the number 3 is, so long as you can still add 1 to it and get 4. More generally, for a term to be defined by the axioms of the system of which it is a part, the elements which satisfy those axioms must not themselves have any use within the system besides relating the term to other terms also defined by the axioms. For example, in the case of points the elements will be either dots on a piece of paper or coordinates (x, y), and in the case of straight lines the elements will be either lines on the paper or the ratio $(u : v : w)$. The elements of points relate to the elements of straight lines in a way which satisfies Hilbert's axioms - the dots are positioned on top of lines to represent the relation *p lies on q*, or the coordinates are such that $ux + vy + w = 0$, and from this it follows that if AB=a and AC=a, where B≠C, BC=a, and so on. But aside from the way in which the elements satisfy the axioms, they have no properties relevant to the system of geometry - it doesn't matter, for example, what colour you have drawn the lines in, or whether you've expressed the numbers in base 12 as opposed to base 10. The only properties of lines, or numbers, which concern us, are those that satisfy the axioms.

So, if implicit definition is to be extended to terms in scientific language, as Schlick believes it can be, we need it to be the case that:
(a) the terms appear in the axioms of the language of science and
(b) the elements which satisfy the axioms of the language have no properties with any use within the language of science other than those that relate them to other elements.

1.2.3 The Axioms of the Language of Science

The question now arises: What, exactly, are the axioms of the system of science (which is not a separate question from: what *is* the system of science)? Don Howard (1993) argues that the answer to this question changes between the first and second editions of AE, although all of the 1918 passages he uses to back up his interpretation of the first edition are retained in the second and Schlick himself refers back to those passages when he clarifies his position in the newly-added §11. Whilst I'm not convinced that Schlick's position changed, looking at both of the views identified by Howard will expose some of the issues at stake.

A scientific theory is a set of judgements (again, Schlick did not switch to talking about statements or sentences until later), and "the *intellectual* labour of science … consists in *inferring*, that is, in deducing new judgements from old ones" (AE, p.33, original emphasis). The apparent difference between editions occurs in the distinction between judgements which were to be accepted into the scientific theory as purely definitional, and judgements which asserted something about the world ("knowledge claims"). In the first edition, Howard says, Schlick did not think that there was an intrinsic difference between definitions and knowledge claims. Schlick points out, for example, that in mathematics there will be more than one set of statements which could be taken as the axioms (AE, p.46), e.g. Peano arithmetic contains as an axiom "there is no natural number n such that $S(n) = 0$", and as a consequence of this along with the other axioms we can prove that there is no natural number n such that $S(S(n)) = S(0)$, but we could still prove all the truths of Peano arithmetic if instead we took "there is no natural number n such that $S(S(n)) = S(0)$" as our axiom and proved "there is no natural number n such that $S(n) = 0$"

from that. Once we accept Hilbert's method of defining mathematical concepts by the axioms in which they occur, and once we see that we can have two systems with different axioms but in which all of the same judgements are true, there is no longer any clear divide between knowledge claims and definitions in mathematics.

Schlick does not appear to acknowledge at this point that there are important differences between axiom choice in mathematics and definitional choices in the factual sciences. It's certainly true that for any set of theorems there will be a number of different possible sets of axioms from which we could derive that set, but the question that Schlick is really addressing is what that set of theorems itself should be. The system of arithmetic is the same, whether we take the Peano axioms or some other appropriate set, and it's then an open question whether or not anything in the world can be made to fit the system (i.e. whether we can pick out something in the world to count as "zero" and define a successor function from it without end or repetition). If we find a realisation of the system then our prior work with arithmetic will tell us things about the realisation, but even if we never find a realisation it remains true that the system can be axiomatized in various ways. Science, on the other hand, starts with something that exists and then looks for the rules that describe it. If electrons can't be assigned a place in the formal system of electromagnetics, then the system won't tell us anything about the behaviour of electrons, and that remains true whichever theorems of the system are taken as its axioms. We'll return to this issue in section 1.2.5 when we look at Carnap's criticism of Schlick's approach.

Traditionally the process of definition in empirical science has been thought to end in concrete definition (i.e. with someone pointing at something), but Howard claims that Schlick thought that a mature scientific theory could eliminate concrete definitions entirely (1993, p.67). This would leave the concepts of empirical science entirely defined by the axioms of the science in the same way as Hilbert's mathematical concepts. It may seem intuitively wrong to say that, for example, *there is snow on the ground* could be part of the very definition of *snow*, or that *snow is a*

state of water might be an empirical fact (and therefore only contingently true), but Schlick argues as follows:

> Every judgement puts a concept into relations with other concepts and designates the fact that this relation exists. If the concept in question is already familiar and defined, then we have an ordinary judgement. If this is not the case, then the concept is to be regarded as having been created by the judgement. The latter thus becomes the definition, which constructs the concept out of its characteristics. It therefore seems quite proper to grant the status of judgements to definitions as well; theoretically, *definitions do not occupy a special position.*
>
> - Schlick, 1918, [1985, p.47], referenced in Howard, 1993, p.66, where emphasis is added by Howard

Take, for example, the judgement *snow is cold*. This puts the concept of *snow* into a relation with the concept of *x is cold*.[6] Specifically, it is judged that any object that falls under the concept of *snow* will have a property corresponding to the concept of *x is cold*. Do we take the judgement *snow is cold* to be part of the definition of the concept of *snow* (or the definition of the concept of *x is cold*)? According to this reading of AE, that depends on the extent of our previous understanding of the concept of *snow*. If we had never used the concept before then it would be a definition: the concept of *snow* would apply to any object that was cold until it was further defined. (It is obviously difficult to imagine circumstances in which we might pick out and name an object solely by its temperature, but other examples demonstrate that it is possible to affix an object concept to a single property concept – consider the concept of *square* which can be applied to any object that has the property associated with the concept *x has four equal sides and*

[6] For the purposes of this discussion I am following Schlick's lead (AE, pp.40-41) in treating "snow" as denoting an object, *snow*, and "cold" as denoting a property, *x is cold*, although I think there are contexts in which it would be appropriate to interpret "snow" as denoting the property *x is snow*.

four right angles.) If we had only a partial understanding of snow, say we knew that it was white and fell from the sky but not what it felt like, then *snow is cold* could become part of the definition, meaning that the concept of *snow* would apply to anything which instantiates *x is cold* as well as *x is white falling stuff* instead of merely the latter, assuming we accepted the judgement as universal and not just true of *this* snow. If we had a complete understanding of the physical structure of snow and how it is formed, then *snow is cold* might not be part of the definition of the concept of *snow* at all - it would just be a consequence of those facts about snow's composition. So, *snow is cold* can sometimes be part of the definition of the concept of *snow*, and sometimes be a judgement about the nature of stuff that falls under this elsewhere-defined concept of *snow*. Hence, there is no clear distinction between definitions and judgements. Schlick took it that the apparent distinction between the two kinds of statement had only arisen in the empirical sciences by accident (AE, p.47) - in this world we first identified snow by its tangible properties and only discovered its structure later, but in another possible world we could have designed a particular molecular structure, called it "snow", and then built it to see what it would look like. The fact that some statements appear to be definitions while some appear to be judgements was just a product of how we have in fact discovered the relations between these concepts, and in a mature scientific theory the distinction would break down. Schlick in fact thinks that at this point the judgements that will be taken as axioms will be the ones that have the fewest primitive terms between them whilst still allowing us to derive the entire theory, arguing:

> At one time, mathematicians regarded as axioms those propositions that seemed especially self-evident; today we do not hesitate in principle to derive such "axioms" partly from less obvious propositions and to look upon these as the axioms (and hence as the definitions of the primitive concepts), if by doing so we can achieve a simplification in the construction and compactness of the system.
>
> - AE, p.46

But this is a purely practical concern and there is no reason why other axioms might not be chosen in principle.

In a newly-added §11 in the second edition, Howard argues that Schlick no longer appears to think that the distinction between definitions and knowledge claims is unclear in the empirical sciences (although he still maintains that it is for "ideal" sciences (i.e. maths and logic)). Despite having earlier argued that which judgements are taken as definitions is an accident of the order of human experience, he now argues that there can never be any ambiguity over whether a judgement is a definition or not. Suppose that we are considering the judgement *all bodies are heavy*. If we take being heavy to be part of the definition of being a body, then *x is heavy* is a characteristic of the concept *x is a body*. If, on the other hand, *all bodies are heavy* is a synthetic knowledge claim then the concept *x is a body* does not have the characteristic *x is heavy*. Thus, although the same words ("all bodies are heavy") are used to express the judgement regardless of whether or not *all bodies are heavy* is an axiom, the concepts involved in forming the judgement are not the same. The judgement which acts as a definition could be expressed as "all extended heavy things are heavy,"[7] whereas the knowledge claim would have to be restated as "all extended things are heavy," which is clearly distinct. Therefore, there is a definite distinction between definitions and knowledge claims, although you won't be able to see it when the theory is written out.

So, there are two views distinguished here: on the first view there is no real distinction between definitions and other judgements, so *snow is made of water*, *snow is cold*, and *snow fell yesterday* all have the same status; on the second account there is a distinction between definitions and empirical sentences, which is found in whether or not the judgement is analytic given the characteristics of the concepts involved. Schlick obviously adheres to the analytic/synthetic distinction and the

[7] This does not, of course, mean that you could do away with the concept *x is a body* by replacing it with *x is an extended, heavy thing* in every case, since if *bodies are heavy* is taken as an axiom then *x is heavy* is defined by *x is a body* just as much as *x is a body* is defined by *x is heavy*.

empirical nature of scientific enquiry, which is going to cause problems for the first view. In §11 he seems to clarify what he intended in the passage emphasised by Howard. He writes "[w]e explained above (§8 [which is the section containing the troublesome passage]) that one and the same (linguistic) sentence can express both a definition and a piece of knowledge," (AE, p.76). The addition of the "(linguistic)" seems to be aimed at avoiding the conclusion that, because the same sentence might at one time express a discovery and at another a definition, there is no true distinction between definitions and empirical judgements.

Concrete definitions still play a role in Schlick's system – the vast majority of concepts are given characteristics concretely. The role of implicit definitions is in fixing the meaning of those concepts which resist concrete definition – the most basic concepts, like time and mass. These are the ones which are most plausibly defined implicitly. It's not much of a stretch to think that, for example, the concepts of electromagnetism are defined by their place in Maxwell's equations, and that they resist further reduction. It's certainly difficult to maintain that the reference of "magnetic flux" is fixed by our sense experience of magnetic flux. Schlick also emphasises that judgements about observed circumstances become hypotheses about the past as soon as we are beyond the instance in which they are formulated (a thought which will stay with Schlick when he proposes founding certainty in affirmations, as we will see in chapter 5), which means there is little value in distinguishing between factual judgements that are known and those that are merely predicted. As such, Schlick claims that there are exactly two distinct kinds of judgement in the system of science: definitions and hypotheses (AE, p.73).

Let's return to the concept of *snow*. We might be introduced to snow as a child by being told "the stuff that fell outside yesterday was snow," and by independently forming on the basis of experience the hypothetical judgement *the stuff that fell outside yesterday was cold and white*. We then form the definitional

judgement *snow is cold white stuff*.[8] This judgement is identical to *cold white stuff is cold white stuff*, and is therefore regarded as analytic. Any judgement containing the concept of *snow* is now reducible to a judgement about stuff which exhibits the property falling under the concept of *x is cold and white*. *x is cold* and *x is white* in turn require definition, but it is clear that their meaning doesn't change by learning that there is something which is cold and white (i.e. snow). As such, *snow is cold white stuff* does not form part of the definition of either *x is cold* or *x is white* – those terms are defined implicitly by their roles in the axioms of thermodynamics and optics respectively.

The axioms of science, therefore, are not all of the analytic judgements in the system, but the judgements which describe the characteristics of those concepts which resist concrete definition – things like Maxwell's equations for electromagnetism, the laws of thermodynamics, and Einstein's field equations for space and time.

When Schlick wrote that it was possible to designate any object by implicit definition (AE, p.70) he was pointing out that the concept of *x is cold and white* can be substituted for some appropriate reference to the laws of thermodynamics etc. Even individual objects (such as the Eifel Tower, as opposed to broad categories of stuff like snow) can be picked out by reference to implicitly-defined terms by using the manifold of spacetime to single out a location. Schlick points out that these definitional judgements are arrived at initially on the basis of experience, but once we have them they no longer rely on experience – part of what we needed to reach our definition of snow was the empirical judgment *the stuff which fell outside yesterday is cold and white* but once we have that definition the fact that snow is cold is a logical truth. Moreover, many things which we might otherwise have to laboriously measure turn out to be logical truths. If, by (simplified) definition, *snow is cold* and *lava is hot*, then we don't have to put snow and lava side by side and touch

[8] This is obviously simplified. The actual definition will say something that distinguishes snow from milk.

them to find out that lava is warmer than snow. The laws of thermodynamics and chemistry, coupled with definitions like *water is the liquid form of H_2O*, give us a range of facts about the temperature and energy level of the oceans which we don't need to find out by empirical investigation. Schlick's own example is that we can, by designating a celestial body by its position at a handful of times, accurately predict its position at any past or future time (AE, p.70).

The extent to which it is possible to make predictions on the basis of definitions depends upon the exactness of the science in question. Schlick appears to distinguish between different sciences with different axiom sets. In particular, the historical sciences are incapable of making predictions because their judgements are largely about matters of fact rather than the defining characteristics of their subject matter (AE, p.78). Archaeologists make a large number of claims about the Kingdoms of Egypt, but these do not reduce to a small set of laws that can tell us how Egyptian society will progress in the future.

All of this means that definitions, and in particular implicit definitions, are crucial to the predictive work of science in Schlick's account. But it doesn't mean that every scientific judgement is a definition. Even if *snow fell yesterday* is accepted into the system, it doesn't take on the characteristics of an analytic truth. In the latter sections of AE, Schlick makes it clear that there are distinct roles for analytic and synthetic judgments in his philosophy: "[Analytic] judgements have nothing whatever to do with *knowledge* of reality, and may therefore be completely separated from it. Their realm is that of thinking, not of being," (AE, p.335, p.285 in the first edition, original emphasis). That is to say, analytic judgements are to do with how we divide up the world for our own comprehension, but say nothing about the way the world actually is. Synthetic judgments, on the other hand, say something beyond what is contained in the concepts involved. "A synthetic judgment … asserts a real interconnection of traits," but "never has the character of a universally valid truth," (AE p.343). Empirical experience therefore still plays a role in Schlick's account, even though implicitly defined concepts are not reduceable to sense

images. The question naturally arises as to how we are to form synthetic judgements about these entities, and the answer comes from Schlick's account of truth.

Schlick's account of truth is, as noted, unwaveringly correspondence-based throughout his career. At the time of AE, his view was that "a judgement which *uniquely designates* a set of facts is called *true*," (AE, p.60, original emphasis) – a view which can also be found in his earliest writing on logic (1910, [1979a, p.94]). [9] The notion of "unique designation" was to be preferred over one like "agreement" because "agreement" is ambiguous and invites the confusion that the judgement must be literally the same as the fact, which is obviously false because judgments and facts are different kinds of thing. As far as truth is concerned, Schlick is prepared to accept *any* system for uniquely associating a judgement with a set of facts. It would be possible, he argues, to name every possible fact individually and say that the true judgements are those names for which the designated fact exists (AE p.66). However, that system would not allow for knowledge. Knowledge requires both a unique designation and a connection to other concepts already understood. That is why concepts must be defined in terms of one another. So, a true judgment within

[9] There is some difference between the earliest Schlick and the Schlick of AE at this point, which may be useful in elucidating his account: in 1910 Schlick thought that truth required a one-one coordination of judgements to fact, whereas in 1918/25 Schlick is content with a many-one coordination. What matters in AE is that every knowledge claim designates only one fact, but one fact may nevertheless be designated by many knowledge claims. It is interesting that Schlick only mentions the softening of his position in brackets and even then only says that the one-one coordination is not "*absolutely* necessary" (AE, p.60, emphasis added), because requiring a one-one coordination would be highly problematic for what Howard claims was Schlick's 1918 view. Definitions do not state facts and are not true or false. If a judgement which is a definition at one time can be a knowledge claim at another then a fact will need to come into existence for the judgement to be coordinated with. Suppose that at t0 the definition of *gold* is *malleable yellow stuff* and it is a fact that all gold has the atomic number 79. Suppose at t1 the definition of *gold* becomes *stuff with the atomic number 79*. At t0 there is one true judgement - *gold has the atomic number 79* which is coordinated with the fact that for all x, x has atomic number 79 if and only if x is yellow and malleable. At t1 there are two true judgements - *gold is yellow* and *gold is malleable*, each of which used to be merely a definition and each of which is correlated to the same fact as *gold has the atomic number 79* used to be. Hence, we can end up with a number of judgements coordinated with the same fact. Nor can we get around this by saying that definitions are, in some sense, true, because there is still only the one relevant fact with which we can coordinate all these judgements: the fact that malleable yellow stuff has atomic number 79. The problem does not arise if, as Schlick implies, all he ever meant by the relativity of the distinction between knowledge claims and definitions was in their linguistic formulations, because for Schlick at this time truth is a property of judgements and not of sentences.

the system of science is one for which the interconnections between the concepts employed are sufficient to uniquely pick out a state of affairs in the world.

How do we get a judgement containing implicitly defined concepts to correspond uniquely with a state of affairs in the world? After all, implicit definitions alone merely characterise an empty structure, so there will certainly be more than one realisation of any sentence that contains such a concept. Schlick's answer is that we can use coordinative definitions and conventions to bring implicitly defined concepts into line with concretely defined terms. For example, we can adopt as a convention that days will always be the same length, and thereby coordinate the concept of time with the concretely determined facts about how long a day is (which will be relative to the durations of other events). Time is not *defined* by the length of a day – its definition is to be found in the axioms of general relativity – but the facts picked out by judgements involving time are to be coordinated with the real world by means of this convention. We could equally well pick something else to designate equal divisions of time – in reality we use the radiation of caesium atoms under certain conditions, but Schlick points out that we could in principle use the Dalai Lama's heartbeat. It would create great practical difficulties – the speed limit would have to be decreased whenever the Dalai Lama was exercising (or else the metre shortened) – but *something* has to be chosen, conventionally, as the basic unit of time, and other measurements will have to be taken relative to that. Adopting a concrete standard as the unit of time gives us a way of using the concept in empirical judgements without requiring that we *reduce* time to that standard. Time is to be coordinated with the length of the day (or caesium radiation), but the length of the day is not what time *is*. And if these conventions cause problems in observation, as for example when Newtonian physics proved incompatible with observation, then either new conventions or new axioms are required.

1.2.4 The Viability of Implicit Definitions

In section 1.2.2 we saw that for implicit definitions to be viable for scientific terms we needed (a) the terms to appear in the axioms of the language of science and

(b) the elements which satisfy the axioms of the language to have no properties with any use within the language of science other than those that relate them to other elements.

We have just seen how (a) is to be satisfied. The terms that we are defining implicitly are precisely those that appear in the axioms of science – terms like “time” and “mass” and “electron”. The question now is whether or not the axioms are *sufficient* to constitute a full definition of the terms therein.

Criterion (b) meets the same kind of objection that we saw Frege raised against Hilbert. One kind of thing that will (hopefully) satisfy the axioms of "electron" is an electron. There are, we may assume for the sake of argument, real subatomic particles which behave in ways that are predictable from the axioms of quantum electrodynamics (henceforth “QED”). The claim, then, is that there are no properties of electrons that contribute anything to the language of science besides their relations to other elements of QED (e.g. photons). In other words, the claim is that there is nothing lost if we use other elements which nevertheless satisfy the axioms – if we write a computer program that simulates the behaviour of electrons, say, then we will be able to say that within the program there are certain lines of code that *are* electrons, just as much as the subatomic particles are electrons relative to the world. If the relations between terms are all that matter then a simulated electron, if it emits simulated photons and so on, isn’t just a model – the simulated electrons have just as much claim to being the denotation of “electron” as actual subatomic particles do.

This claim is certainly in conflict with the common-sense view of language. It seems as though words like "electron" don't just signify a bunch of relations, they signify a particular thing, and any relations between "electron" and "photon" (ignoring ones like *has more letters than*) are merely in virtue of the relations between the subatomic particles which the two words signify. Just as Frege objected that once you start talking about (x, y) you are no longer really talking about points, the common-sense objection runs that once you start talking about computer programs you are no longer talking (directly) about electrons.

Perhaps, however, the same reasoning Benacerraf puts forward on "3" can be brought to bear in this case. Suppose that there are features of electrons that don't appear in the QED model. For example, perhaps we can imagine that some electrons are cube-shaped while others are spherical. If these electrons are experimentally identical – they absorb and emit photons in the same way and travel in the same way – then what does the shape of the electrons add to our use of the word "electron"? If you ask someone to calculate such-and-such probability amplitude, it won't matter if you were working with cube-shaped electrons and they interpret you as asking about spherical electrons because the answer is the same either way. This is analogous to Benacerraf's argument that it doesn't matter what the cardinality of "three" is for you, so long as you can still carry out basic arithmetic. If the function of the axioms of science is to make predictions about experience, aspects of electrons which play no role in the predictions also have no place in the axiom system. If the shape of an electron *does* make a difference somewhere along the line, then it will be time to switch to a new axiom system to make predictions – such is the nature of scientific advancement. Thus (b) is satisfied and the way is clear to defining theoretical terms with the axioms of the language.

1.2.5 Carnap's Objection

In "Eigentliche und Uneigentliche Begriffe" ("Proper and Improper Concepts") (1927), Carnap distinguishes between concepts that are defined concretely and those that are defined implicitly. The former he calls "proper", and the latter "improper". Distinguishing them, he writes:

> Empirical concepts are constituted step by step in the systematic construction (Aufbau) of our world-knowledge. Each empirical concept, as a component of this structure, has a direct connection to reality. In contrast, the improper concepts hang in the air, so to speak, awaiting instructions. They are introduced by an axiom system, but that system doesn't relate directly to anything real. The axioms of this system and the theorems deduced from it don't properly form a theory (as they aren't actually about anything in

> particular), but rather just a theory-schema, an empty framework for possible theories,
> Carnap, R., 1927, translated and referenced in Awodey & Carus, 2001, p.149

Carnap's principle objection here is that a network of implicit definitions doesn't say anything about the world. Just as the Peano axioms define the same structure in this universe or an empty one, the axioms of QED define the same system whether or not there are any such things as electrons. The consequences of this are that, although we can use QED to make predictions about electrons when electrons have been determined to have a suitable structure, this determination itself needs to be made in experience. We can define any number of axiom systems in the abstract, but none of them can be said to define "electron" until we have actually investigated electrons and found out what properties they have. The problem here is related to, but different from, the objection of Frege's that we discussed above, made against Hilbert. Frege had objected that points and lines might have some properties that are not captured in the axioms that are supposed to define them, but we've argued that Benacerraf was right to say that mathematical terms simply have no properties other than those that relate them to other mathematical terms. Carnap's objection here is not that electrons have some existence independent from their relational properties – as we will see, he thinks that structure is sufficient for the definition of any scientific term – but he does worry about whether or not anything in reality instantiates the relational properties stipulated in an implicit definition, and thus argues that for a structure to adequately describe a concept we must understand what it would be for that structure to be found in the world.

Carnap illustrates his point by considering the ways we can apply the Peano axioms in the world. The axioms are relevant any time we have a sequence with a start but no end, no repetitions and no loops. We can define such a sequence along the edge of a desk. We can take one corner as 0, the halfway point to the other corner as 1, halfway again as 2, and so on, getting half way closer to the edge of the table each time but allowing for an infinite number of points to be picked out in sequence.

With this definition in place we can now use the predetermined results of arithmetic to calculate things about the edge of the desk. If a snail is crawling along the desk at a rate of half-way-to-the-end per minute, we know that it will reach the point designated by "10" at the end of the tenth minute. But creating this mapping from table edge to integers doesn't mean that points on the table edge *are* integers. The Peano axioms don't now come to define the table. The concepts of Peano arithmetic are useful when such a sequence is found, but on their own they don't denote anything in the real world. Nothing intrinsically *is* a number because the properties of numbers are not the kinds of thing that objects in reality instantiate, and likewise, Carnap says, no real thing can be defined by an abstract structure from implicit definitions. In contrast, for proper concepts we don't need to say "such-and-such falls under this concept *if we interpret it in a certain way*". We don't say "Fido is a dog if we can find a way to map his DNA onto our Dog Concept DNA" – we say "we investigated Fido, found that he possessed the characteristics of a dog, and therefore have determined that he is a dog".

Carnap's position was that the only way a concept could really be defined was by some link to empirical experience. If an object can be shown to have the properties attributed to a variable in the axiom system, suitably interpreted, then we have a realisation of that system and can use the axioms to make predictions. If we can show that electrons have the right properties, we can use QED to predict their behaviour. But it is essential that *showing that they have the right properties* occurs, and this can only be done by experience. However, Schlick's arguments against the use of intuition to ground scientific concepts still stand, and so the role of experience can't depend on the intuitive content of that experience.

Carnap did suggest an alternative to implicit definitions that would not result in Intuition re-entering the picture as a component of knowledge. In *Der Logische Aufbau Der Welt* (1928, [1967, §13-15]; henceforth "the *Aufbau*") (which was in development and known to Schlick several years before publication), Carnap develops the notion of structure descriptions. Where implicit definitions serve to define *properties* in terms of networks of relations between one another, structure

descriptions pick out individual *objects* by the relations they hold to one another. Instead of definition by axioms, relations are described in totality by listing the objects between which they hold. This gives an extensionally-adequate way of picking out objects uniquely unless they are equivalent in all respects that science can distinguish, and if that is the case then science must treat those objects as identical. This picking out was to be done in experience by looking at the relations that actually hold between objects and determining how the network of relations in the world could be made isomorphic with the network of relations in the structure description.

The advance that structure descriptions introduce which takes them beyond implicit definitions is a way of bringing experience back into the picture without thereby introducing the imprecision that so bothered Schlick. Carnap's complaint was that implicit definitions don't have an appropriate link to empirical facts, but Schlick's arguments about the imprecision of intuitive content still carry the same force that they did before. The response, therefore, is not to reintroduce intuitive content to the picture of scientific knowledge, but rather to purge what little of it remains in our account of meaning. After Carnap's criticisms are taken on board, even terms that we might think could be defined straightforwardly by ostension, such as "blue", will in fact be picked out by structure.

Schlick appears to endorse Carnap's account shortly after publishing the second edition of AE, writing "Cf the acute and irrefutable remarks of Rudolf Carnap in his forthcoming work, *Der Logische Aufbau Der Welt*, where he shows that all scientific judgements must confine themselves to purely structural assertions," (1926, [1979b, p.111]; Uebel, 2007, pp.75-76), and, as we noted before, the term "implicit definitions" does not appear in his published work after Carnap's criticisms emerged. Nevertheless, many of the commitments which drew Schlick towards implicit definitions remain in place through his later philosophy. In particular, the distinction between knowledge and acquaintance, along with the resistance to the idea that perceptual images themselves enter into knowledge, remains in Schlick's philosophy throughout his life.

Carnap's objection convinced Schlick of the need to more closely tie experience to the system of scientific concepts, but the problem wasn't that formalisation had removed intuitive experience from the picture – it was that effecting a separation between the everyday language and the language of science made concepts in the language of science only ambiguously related to the things we find in everyday life. The solution to this was not to be found by relating the concepts of science back to intuitive experience, but rather by unifying the languages and ensuring that experience could be understood in a structural way rather than by ostensive definition.

1.3 Conclusion

We have seen that Schlick's considerations of knowledge led him to the idea that it always involved putting at least two concepts into a relation with one another, and that intuition, involving as it does only one object, therefore cannot provide knowledge. We saw that scientific knowledge requires absolutely precise concepts, which means that the concepts must be defined by other concepts rather than with reference to experience, which is never clear enough. And we saw that Schlick's attempt to define concepts implicitly was to fail to define them as concepts at all, at least in the eyes of Carnap. Implicitly defined concepts are left as mere variables in a logical system, and if they are to say anything at all then we need to determine in experience whether anything actually has the properties of the implicitly defined concept. But this determination in experience cannot involve the reintroduction of intuition, so if experience is to be linked to the purely-structural concepts of science then experience itself must be describable in purely structural terms. Carnap offers a method of achieving this goal, which appears to be endorsed by Schlick shortly after the second edition of AE, and which we will explore further in chapter 4. Before that, however, we will introduce the work of Wittgenstein and look at the influence that it had on Schlick's development. We will see that the position illustrated by Wittgenstein would have further moved Schlick to adopt a view of language that was entirely structuralist, from the everyday concepts of immediate experience to the

purely theoretical concepts of science, which was the view he would hold at the time of the protocol sentences debate towards which we are moving.

2. The Influence of Wittgenstein

2.0 Introduction

Wittgenstein's *Tractatus Logico Philosophicus* (1922; henceforth "TLP") had a profound influence on Schlick's philosophy. Some commentators consider this influence to mark a clear division between phases of Schlick's philosophy (e.g. Waismann, 1938, [1979b, p.xvi]; Feigl & Blumberg, 1974, p.xvi), others see Schlick's thought as essentially unified in pursuit of solid foundations for science (e.g. Geymonat, 1985, pp.273-274), but even those in the latter group agree that there were substantive changes in Schlick's position.[10] In this chapter I shall draw out exactly how Schlick adopted Wittgenstein's views with regard to how language is able to express facts about the world. I will show that there are important ways in which Schlick had anticipated the ideas of the TLP in AE, but also that these ideas underwent substantial development by the time of F&C.

In the first section I will discuss Wittgenstein's logical atomism and picture theory as a precursor to investigating the way in which these theories impacted Schlick. In the second section I will highlight a number of key areas in which it is clear that Schlick's thought developed as a result of Wittgenstein's influence. In the third section I will refer back to the work of the previous chapter and we will see how the TLP provides a way forward for Schlick out of implicit definitions.

2.1 The *Tractatus Logico-Philosophicus*

In this section I will outline Wittgenstein's logical atomism and picture theory - two

[10] As we noted in the introduction (page xviii), the question of exactly when this influence was exerted is a matter open for debate.

ideas which were central to the TLP and which show their influence clearly in Schlick's later work.

2.1.1 Logical Atomism

The world is everything that is the case (TLP: 1) and substance is what exists independently of what is the case (TLP: 2.024). The world is the totality of facts (TLP: 1.1) and facts are determinate combinations of objects (TLP: 2.01). Objects form the substance of the world (TLP: 2.021).

For Wittgenstein objects are things that exist necessarily, and different possible worlds are different possible arrangements of objects. His argument for the necessary existence of objects is obscure, but his conclusion is clearly stated. For our purposes, nothing turns on the interpretation of the argument and so I shall not make the attempt.

Given that objects unquestionably exist they have to be simple - they cannot be composed of anything else. This is because if they were composed of something else then their existence would be not necessary, but contingent upon those somethings else being so combined (White, 2006, p.41); "Every statement about complexes can be analysed into a statement about their constituent parts" (TLP: 2.0201). If we take, for the sake of illustration, a table to be a complex consisting of four legs and a flat surface then the statement "there is a table" could just as well be expressed as "there are four legs attached to a flat surface", which could itself be broken down into a sentence about individual particles of wood. The statement about the table could not be true if not for the statements about the parts. Note that this shouldn't be taken to imply that particles or atoms are actually objects in the sense that Wittgenstein is talking about. Perhaps the existence of an atom consists in the existence of a complex made up of, say, space and time and mass. It's certainly easier to accept time as having necessary existence than any individual atom. Wittgenstein resists providing any example of what an object might be, limiting himself to explaining their formal role in his philosophy.

Objects are said to have both internal and external properties (TLP: 2.01231). Internal properties are properties which an object *must* have if it exists,

and internal relations as relations which *must* hold between objects if they exist (F&C, p.294). External properties and relations, on the other hand, are contingent on the state of the world. So, for example, imagine x is a spatial object, which is to say it is an object which is located somewhere in space. It is an internal property of x that there is *some* point in space at which it is located, but it is an external property of x that it is located at the particular point in space where it in fact is. We could not conceive of a spatial object not being somewhere in space, but we could conceive of it being somewhere else. If we think of colours as objects (and they seem like good candidates) then, for example, the internal properties of purple would include something like *purple is between red and blue*, and *purple is a colour*, whereas the external properties would include *this plumb is purple*. Wittgenstein describes internal properties as being the *possibilities* of the object's occurrence in atomic facts (TLP: 2.0123), while external properties are the object's *actual* occurrence in atomic facts.

To know an object is to know its internal properties - not its external properties (TLP: 2.01231). As we have seen, it is an internal property of an object that it could have external properties of a particular kind, so knowing the internal properties of an object tells us what kinds of situation that object can be put into. If we know what a speck in the visual field is then we know how it can be related to other specks and we know that it makes no sense to try to relate it to a tone (along the most obvious dimensions, although it might still be related by terms like "more/less extreme than" or "later than"). This is what Wittgenstein means when he writes "Objects contain the possibility of all states of affairs" (TLP: 2.014) - if you know the object, then you know all the possible external properties of that object, so, if you know all of the objects, then you know all of the possible ways the world can be.

2.1.2 The Picture Theory

Explaining how the names in a proposition hang together to express a fact is the goal of Wittgenstein's Picture Theory. A picture, in this sense, is something in which the various parts stand in for the parts of the thing being pictured and their

arrangement matches the arrangement of the pictured situation. A map is a picture of the terrain in a particular area, in which the various shapes and colours represent features of the landscape like woods and rivers, and in which the spatial arrangement of the shapes is supposed to mirror the spatial arrangement of the features they represent in a scaled-down, two-dimensional way. In contrast to the shapes and colours which designate things,[11] the contour lines on the map play a role in designating a relation. The relation *x and y are on opposing sides of a contour line* on a map represents the relation *x and y are at different heights* in the real world. It's important that things are represented by things while relations are represented by relations. As we will see below, the elements of a picture have to be of the same type (thing, one-place relation, two-place relation, etc.) as whatever they represent.

Propositions picture the facts of reality in much the same way as the map, but, since writing is all spaced linearly, relations between things have to be indicated by words. The fact that A is to the left of B, for example, is not expressed by literally putting 'A' to the left of 'B' (i.e. "AB") - that would leave us with a language which could only express whether something was to the left or right of another thing (i.e. for any two things *x* and *y* we would have to choose between saying "*xy*" or "*yx*"). Instead we put the names in the more complex relation of 'A' being to the left of 'B' with the words 'is to the left of' between them, producing "A is to the left of B". We could also express it by putting 'B' to the left of 'A' with the words 'is to the right of' between them (i.e. "B is to the right of A"), which is an equivalent proposition in English. By putting the right words between 'A' and 'B' we can express any number of relations which A might bear to B, be they spatial, temporal, visible, audible, familial, pedagogical, or otherwise.

"In the proposition, the name represents the object," (TLP: 3.22). How is it that we set up this naming relation between symbols and objects? How do elements

[11] Here I'm using "things" to refer to stuff which we might in everyday language call "objects", like trees and fields and pencils, but not relations.

of a picture come to represent objects in the world? For Wittgenstein, the answer seems to be something like that process which we saw Schlick reject in the last chapter - ostensive definition. But here it's important to remember that objects, for Wittgenstein, are simple things; things which cannot be reduced to anything else and which cannot be explained in any other terms. He writes that a "proposition can only say *how* a thing is, not *what* it is," (TLP: 3.221) which is to say that a proposition can only put an object into a relation with other objects. A proposition is an arrangement of elements and says that the objects represented by those elements are in that same arrangement in the world. Telling us where an object is in relation to other objects doesn't do anything to explain the object itself. Explaining objects is therefore beyond the limits of what a proposition is capable of. And, therefore, naming Tractarian objects cannot be done explicitly using already-known names. We cannot use explicit definitions and say "the object which we'll call 'N' is the one which φ".[12] We need a different approach.

According to Wittgenstein, the meanings of names are learned by way of "elucidations" (*Erläuterungen*), which he only briefly and cryptically explains:

> The meanings [or 'references'] of primitive signs can be explained [or 'made clear'] by elucidations. Elucidations are propositions which contain the primitive signs. They can, therefore, only be understood when the meanings of these signs are already known *(kennen)* [or 'only be understood when one is acquainted with the reference of these signs'].
> - TLP: 3.263, alternative translations are Anscombe's, referenced in Helme, 1979, p.323.

Elucidations are not quite ostensive definitions, because they are propositions rather than actions (or action-proposition pairs) and because they are

---

[12] At least, not to *define* "N". A proposition like this might be used to pick N out so that we might become acquainted with it.

not really definitions. In a definition the word which is being defined cannot be used, it can only be mentioned. To define "cat" we would say " 'cat' means small, fluffy,…", but not "cats are small, fluffy,…" - the former is a stipulation, the latter an assertion. Elucidations are more like the latter example. Hacker suggests "This is A" would be an example of the kind of proposition Wittgenstein had in mind (1986, p.77), and Wittgenstein's later remarks seem to confirm this (1930, [1975, §6]). There is a tension between the dual definitional and propositional roles that elucidations are supposed to fulfil - whilst we could learn something about the meaning of "cat" from a proposition like "cats are small, fluffy,…", we can only do so if we assume that the proposition is true. If we take it to be an assertion and try to establish its truth value without any other source for the meaning of "cat" we are going to be in trouble. Wittgenstein seems to have realised this by 1930, and his position in the *Philosophical Remarks* moves in some ways closer to Schlick's earlier position, just as Schlick was moving closer to Wittgenstein's. We will explore this further in section 3.

Returning to the above example - "A is to the left of B" - there is some question as to whether for Wittgenstein the words "is to the left of" refer to the relation between A and B or whether they are merely part of the relation between "A" and "B". In the first case there would be three objects comprising the fact that A is to the left of B - A, B, and the relation between them - and there would be correspondingly three parts of the proposition. In the second case A and B would be the only objects. Wittgenstein wrote:

> We must not say, "The complex sign '*aRb*' says '*a* stands in relation *R* to *b*'"; but we must say, "*That* '*a*' stands in a certain relation to '*b*' says *that aRb*".
> - TLP: 3.1432, original emphasis

Which many have taken to imply that there are only two parts to the proposition "*aRb*" - '*a*' and '*b*' - which stand in the relation of having an '*R*' between them. Why can we not say that "*aRb*" says that *a* stands in relation *R* to *b* if not because "*R*" is not truly a part of the proposition? On the other hand, relations

appear to be good candidates for being objects - they are generally not reducible to anything else, and it surely wouldn't make sense to imagine a world in which certain relations did not exist, so they meet the necessity requirement. They would also seem to be the kind of things which get named. Words are attached to relations in much the same arbitrary fashion as they are attached to colours or tones.

Thomas Ricketts (1996, pp.69-73) makes the case for the two-part interpretation. He points out that for any relation *R* there is a converse relation *S* such that *aRb* iff *bSa*. For example, the converse of the relation *x is the child of y* is *x is the parent of y*, so *a* is the child of *b* iff *b* is the parent of *a*. Both *R* and *S* describe the same situation, but they are not straightforwardly interchangeable - obviously if *a* is the child of *b* then *a* is not also the parent of *b*. Two names for the same object are always interchangeable so far as truth conditions are concerned - if it's true that Bruce Wayne lives in a mansion then it must be true that Batman lives in a mansion - so *S* and *R* cannot both name the same relation, yet there is only one relation being described by both. If *aRb* says that *a* is the child of *b* then *bSa* also says that *a* is the child of *b*. The fact that we have two different ways of describing the relation makes no difference, since it's clear the language could still describe everything about the world with only one. The problem disappears if we accept that "*R*" and "*S*" are not names and *R* and *S* are not objects.

Stenius (1964, pp.62-63) presents an argument for a three-part interpretation instead. First, consider the (possibly atomic) fact *A is red*. As a fact, this must be a combination of objects (TLP: 2.01). One of the objects is A. The property *x is red* must be another one. There's nothing else for the second object to be. In a picture of a fact, names stand in for objects and, one way or another, they match the configuration of the objects in the world. "A" will stand in for A. What will stand in for *x is red*? Presumably it's "is red". Hence "is red" is a name. And if both objects and one-place predicates are names, it seems awfully arbitrary to draw the line there and say that two-place predicates (i.e. relations) are not names too. Additionally, Wittgenstein writes "in the atomic fact objects hang one in another, like the links of a chain," (TLP: 2.03) which strongly gives the impression of a direct

connection between objects in a fact (and hence between names in a proposition) rather than a connection mediated by unnamed relations.

Stenius' argument seems fairly conclusive, and Proops (2013, §1) gives historical reasons for thinking that Wittgenstein intended the three-part interpretation, but what then are we to make of the fact that two non-interchangeable words can describe the same situation? The answer is that, on Stenius' interpretation, relations are not represented by names in propositions. Stenius (1964, pp.130-132) argues that if we take the sentence "the moon is smaller than the Earth" as picturing the fact that the moon is smaller than the earth with three names then the picture will have failed to be isomorphic with the fact. "The moon" names the moon, and "the Earth" names the Earth, but "is smaller than" is not a name for the smaller-than relation. The reason is that the elements of a picture should always represent something of the same category as themselves. In "the moon is smaller than the earth" "the moon" is a linguistic object - it is a name made up of the first two words - and as such it represents an object - the moon. Similarly, "the Earth" is an object - the name made up of the last two words - and represents an object - the Earth. The last part of the fact not yet represented is a two-place relation between the moon and the Earth, so the element of the picture which represents it must be a two-place relation. In this case it is the two-place relation between "the moon" and "the Earth" - that of being on their respective sides of the words "is smaller than". "is smaller than" is not a name, and is not even an element of the picture on its own, but is indicative of the relation between the two names in the proposition. The two names and the two-place relation are the three elements of the picture and they correspond to two objects and a two-place relation between them. This interpretation accounts for the fact that relations or predicates must be indicated by *something* in a proposition without requiring that the element in question be a proper name.

A picture of a fact, then, has two essential aspects. Firstly, it has a number of elements, each of which represents, by arbitrary convention, exactly one part of the fact it pictures. Secondly, it has a configuration. A picture is itself a fact - each of the

elements of the picture is an object and they are arranged in a particular way. Each element represents the same kind of object as it itself is, so simple objects represent simple objects, one-place predicates represent one-place predicates, and so on. The fact that two elements of the picture are related in a particular way says that the objects they represent are related in the same way, and the picture is true iff the objects are in fact so related.

2.2 Parallels and Divergences

In this section we will explore a few of the ways in which Wittgenstein's views clearly influenced Schlick's philosophy. There are three views to compare at this point: Schlick's position as it stood at the time of AE, which we looked at in chapter 1; Wittgenstein's position which we have just exposited; and the position that Schlick was to adopt as a result, which will be the subject of the next three chapters. In the interest of being able to carry out this comparison, the first subsection below will be a brief outline of the position which Schlick developed between about 1928 and 1936. This position is primarily drawn from Schlick's "Form and Content" lectures of 1932, which are where he gives the fullest exposition of his view, but for evidence of how and why his thinking is changed in light of the *Tractatus* it will also be illuminating to look at what Schlick wrote about it in "The Turning-Point in Philosophy" in which he credits Wittgenstein with precipitating a change in philosophy that would be "altogether final" (1930a, [1979b, p.155]). A much more detailed critical exposition will be forthcoming in chapter 3.

2.2.1 Schlick's Later Position

In F&C, Schlick lays out a position which centres around the distinction between logical form and intuitive content, and he makes it clear that only logical form can be known or communicated. The reasoning is that there must be something which is in common between facts and the sentences which express those facts, or between sentences in different languages which express the same fact. This thing that the sentences have in common would today be called the

"proposition" which each sentence expresses, but Schlick described it as logical form.[13] It is something which persists when the elements specific to the fact or the utterance (i.e. the substance of the world or the words on a page) are abstracted away. All that is left in this case is the Ramsified form of a sentence[14] – a logical form. Schlick's view is that expression is what happens when the logical form of a fact is translated into a new material. That which is left behind by this procedure is the content – that which originally instantiated the logical form. In "The Turning-Point in Philosophy", Schlick puts it like this:

> [P]ossible modes of presentation, so long as they really express the same piece of knowledge, must … have something in common, and this common factor is their logical form. … [E]verything else therein is inessential and contingent material of expression, no different, say, from the ink we use to write down a sentence.
> - Schlick, M., 1930a, [1979b] "The Turning Point in Philosophy", p.156

Qualia, and whatever else metaphysicians might try to express that is beyond the realms of scientific inquiry, are inexpressible, and therefore any work which professes to express them is nonsense. What people are trying to talk about with qualia is the material which makes up consciousness – it is the content which, in the mind, carries logical form, but which is *of necessity* left behind when those thoughts are expressed, just as the ink on a page is left behind when you read the words.

Facts are expressed by sentences, and the relation of Expression stands in contrast to that of Representation. Words represent objects in a sentence, and the

---

[13] Talk of logical form also does something which talk of propositions does not: it explains the correspondence between sentence and fact, not merely sentence and sentence.

[14] The Ramsified form of a sentence is that sentence with constants (particularly theoretical terms) replaced by variables of an appropriate type, which are then bound by similarly-typed existential quantifiers. So, for example, "Pa & Rab" will be Ramsified as something like "∃X ∃Y ∃m ∃n (Ym & Xmn)". For our purposes, Ramsey sentences provide a good way of bringing out the logical form of a sentence without getting mixed up with any of the intuitive meaning that we might attach to a term.

word “represent” is here used as English for the German term “*vertreten*”, which might also be translated as something like “stand in for” or “deputize for”. Words are the representatives of objects in a sentence, and the arrangement of those words is supposed to correspond with the arrangement of those objects being represented. Already at this point the relationship to the TLP should be becoming clear. Learning a language consists, according to Schlick, of learning both the vocabulary of the language (which words represent which objects) and the language’s syntactic rules (how we can fit those words together to represent certain states of affairs).

To support this view, Schlick needed an account of the role of syntax in language. Syntactical conventions establish the logical relations which are pictured by a sentence and the logical relations which hold between sentences. They are the rules which tell us that “the uniformly-coloured region is both red and green” is necessarily false, and that if “A implies B” and “B implies C” are both true then so must be “A implies C”. These are the rules which sentences cannot violate without creating nonsense, although they cannot be expressed themselves because they do not *say* anything about the world. They merely show themselves in the possible uses of language (F&C, p.295).

To illustrate the account, consider:

(S) The sky is blue.

S is made of black marks on paper (or on a screen, depending on your viewing choices) arranged spatially, and it expresses the fact that the sky is blue. S is not made of the same material as the fact which it describes - the sky itself is obviously not an arrangement of black marks on paper. The letters <T, h, e, s, k, y,> together appear to represent the sky, and the letters <i, s, b, l, u, e,> seem to indicate an external relation of similarity between the sky and other blue objects (or a relation of a particular kind of difference between the sky and non-blue objects) as dictated by the semantic rules of English. The syntactic rules of English tell us that when "is blue" follows "The sky" the writer is expressing the fact that the property of blueness

belongs to the object which is the sky. The structure of blueness (i.e. as a property which is possessed by all and only those objects whose colour is between green and purple) is mirrored by the rules for using blue in a sentence (i.e. it must be ascribed to an object and the objects to which it is ascribed cannot be said to possess a colour between red and yellow (or at any other non-blue position on the spectrum)). And an aspect of the structure of the sky (i.e. having an external relation which places it between grass and amethyst stones along a certain dimension[15]) is mirrored by the way the words "the sky" are slotted into the object position of the "is blue" predicate. The fact that the sky fits between grass and amethyst along the colour dimension is what makes S true when interpreted according to the rules of English. S would be false if it was interpreted with the rules of another language where, for example, "is blue" represented the relation of being between red and yellow, so the choice of language is important. Equally important are the facts of the world – if we imagine a possible world which is identical to our own with regard to how amethysts and grass and all other objects except the sky relate to each other with respect to colour, but in which the sky does not fit between amethysts and grass along that dimension, then we are imagining a world in which S is inconsistent with the other reports of colour which are true in this world. (I hesitate to say that S is false in this world, because we might want to say that S is the only colour report which is true in both this world and that world, and every other object has shifted some way along the colour spectrum. Both are perfectly valid ways of describing that scenario.)

With this picture of Schlick's later philosophy laid out, we can clearly see the development of Schlick's thought from AE through the catalyst of the TLP.

2.2.2 Picture Theory

In AE, Schlick writes "a judgment pictures the nature of what is judged as little as a musical note pictures a tone, or the name of a man pictures his personality," (AE, p.61), arguing simplistically that "in the judgement 'The chair is to the right of

[15] The meaning of "blue" is obviously not tied to the existence of grass and amethysts, but it is tied to the existence of *some* coloured objects, otherwise there would be nothing to relate it to.

the table', the concept 'chair' is not placed to the right of the concept 'table'," (AE, p.61). It is interesting to note that he wrote these remarks at least three years before Wittgenstein's first publication of the TLP, and perhaps (by Oberdan's timeline) as many as ten years before he read it himself.

In his later philosophy Schlick sets out on a more careful enquiry to establish how it is that by seeing or hearing a sequence of signs (e.g. black marks in a newspaper) which we have never seen or heard before we can come to know facts about things we have never experienced (e.g. a volcanic eruption on a distant island). The first notion he introduces for this purpose is that of expression. "We say that one fact (the arrangement of little black marks) *expresses* the other (the eruption of the volcano), so that the relation between them is called Expression," (F&C, p.286).

There is an immediately apparent parallel between the later Schlick and Wittgenstein in the explicit talk of propositions as facts. Expression is not a relation between two objects, or between a fact and an object - it is a relation that holds between two facts just as the relation of Picturing was for Wittgenstein. Indeed, in his 1930 work Schlick directly credits the understanding of expression to Wittgenstein, writing "the decisive turning point was first reached by Ludwig Wittgenstein … The road to clarity … starts from the fact that every item of knowledge is an expression or presentation. It expresses the state of affairs known in it," (1930a, pp.155-156)

Schlick next introduces the notion of *representation*: "a sort of correspondence between two things which we establish arbitrarily by agreeing that the one shall stand for the other, shall replace it in some given context, serve as a sign or symbol for the other, or, in short, *signify* it," (F&C, p.287). In this case, the relation is between two *things*, not facts. The notion of representation here is exactly the same as that between the elements of a picture and the objects comprising the pictured fact. The relation between the elements of a picture and the pictured objects is entirely arbitrary and the elements of the picture, in a sense, replace the

original objects in the context of the picture, filling in the gaps in the logical form which the objects themselves occupy in reality.

In spite of his clear objection to the picture theory at the time of AE, there are some elements of Schlick's early account of meaning which survive the transformation through Wittgenstein. As we noted on pages 21-22, Schlick had always maintained that unique designation was a feature of truth. In AE, he thought that it was the *only* feature of truth, and the interconnection of concepts was just a means to that end – the fact that concepts appear in a large enough number of different judgements is sufficient to uniquely pick out the fact denoted by any given judgement, which gives us all that we need to say that the judgement in question is true. This interconnection of judgements is also necessary in order for these judgements to constitute knowledge. So, the fact that judgements *have* a structure, which involves words that denote the same objects/properties appearing in multiple propositions and being put into relations with one another, is there from Schlick's earlier philosophy. But the claim that structure has to do with meaning directly is new. At the time of F&C, Schlick no longer thinks that we can name facts directly (as we saw he did on page 22). Facts are not the kind of thing that get names (like people or musical tones) – objects get names and facts are compounds of objects, so facts are denoted by compounds of names. Representation and Expression are rightly to be distinguished, and it's a distinction that only emerges after Schlick adopts some of Wittgenstein's ideas from the TLP.

So, the interconnection of judgements/propositions emerges in Schlick's later philosophy. The exact manner in which propositions relate to one another – the grammar of the language – does, however, introduce a divergence between Schlick's account and that of the TLP.

2.2.3 Grammar

The first difference between Schlick and the *Tractatus* appears over what one needs to know in order to be able to speak a language. Schlick writes that learning the vocabulary of a language is insufficient - "We must study its *grammar* also," (F&C, p.287). The grammar of a language is "everything which can be laid down

before the language is employed" (Waismann, 1938, [1979b], p.xxvi), including rules for mathematics, rules such as "no uniformly coloured surface is both red and green", and definitions of any words which can be defined. For Wittgenstein, on the other hand, to know an object is to know its logical form (TLP: 2.01231), which includes all of the ways the object can be combined with other objects. Since propositions picture facts by matching their structure, and since all possible structures are known once the objects we have names for are known, to know a language's vocabulary *is* to know its grammar for Wittgenstein at the time of the *Tractatus*.

The difference here is not necessarily down to Schlick, however. In "Remarks on Logical Form" (1929) Wittgenstein attempted to use something like the notion of grammar to solve the colour exclusion problem. The problem is that it seems like a logical truth (or, at the very least, an *a priori* truth, which amounts to the same thing for Wittgenstein (TLP: 6.37)) that a uniformly coloured object cannot be both red and green, but there is no obvious formal contradiction in the sentence "this is red and green" (where "this" should be taken to refer to a single point at a particular time or to an object which we have already established to be uniformly coloured at the time/s in question). The sentence has the same form as "this is red and big", which can obviously be true under the right circumstances. In 1929 Wittgenstein proposed the solution that some properties exclude each other by their nature, the rules of language forbidding us from ascribing more than one of them to a particular object at a particular time. Whether or not this solves the problem, it could very well have provided the inspiration for Schlick's concept of grammar which first arises in 1930 alongside a reference to Wittgenstein's paper (Schlick, 1930b, [1979b, p.169]).

The manner in which expression and representation are linked for Schlick is very much the same as how the two parts of Wittgenstein's picture theory come together. An expression matches the structure of a fact in a new material (F&C, p.300). The rules for using words in a language mirror the logical form of the objects those words represent, and the possibilities for meaningfully combining words mirror the possibilities for objects to combine in the world.

2.2.4 Form and Concept

There's a notable distinction between early-Schlick's concepts and Wittgenstein's logical form. A concept is, more or less, a set of properties which a thing may or may not have. A logical form is the logical form *of* a particular object. We could imagine a concept which didn't signify anything (e.g. Unicorn) but logical form presupposes an object. Even when Schlick switches from talking about concepts to logical forms, he is clearly talking about the forms which objects must have *if* they exist - he does not presuppose that they do.

There is an issue here in conflating "object" in the Tractatus with "object" in AE. Schlick's objects are anything we can talk about (AE, pp.20-21), where Wittgenstein's are strictly defined simple things. Schlick's objects include tables and chairs. Wittgenstein's do not. Wittgenstein's objects will fall under Schlickian concepts, because we name objects and in the system of science all words designate concepts, but lots of concepts will not be concepts of objects. Anything which can be defined in the way that most of Schlick's concepts are defined will not be a Tractarian object, because if it's defined in terms of simpler things then it's the simpler things which are better candidates for being objects. The Tractarian objects will then be those primitive things that resist further reduction and that can therefore only be defined implicitly. That is, for Schlick, to understand the concept of a Tractarian object is to understand its place in the axiom system of science, which is the same as merely knowing its logical relationship to other objects. Whilst Schlick does not make the same ontological claims as Wittgenstein regarding simple objects, it seems like the concept of *blue* for Schlick is nothing more than the various ways in which blue can relate to other objects, and the inferences we can make from propositions with "blue" in them. This leaves a much smaller gap between form and concept than first appearances suggested – the picture here is of a Schlick who anticipated some of Wittgenstein's ideas in AE and retained them by a different name in F&C. And since all other concepts are defined in terms of the primitive concepts, they too will ultimately be nothing more than a system of logical relations.

2.2.5 Knowing

For early-Schlick, to know an object is to be able to correctly say of it that it falls under particular concepts (AE: pp.7-8). For Wittgenstein, to know an object is to know its internal qualities, not its external ones (TLP: 2.01231). The difference here may stem from the fact that the two aren't really talking about the same process. In the passages referenced, "knowledge" is being used as the translation for "*erkennen*" in Schlick's case and "*kennen*" in the case of Wittgenstein. Schlick is talking more about the act of recognition, (e.g. to recognise that this, Fido, has the characteristics of a dog) where Wittgenstein is talking about becoming acquainted with the object. As such we could say that the two are not in any kind of disagreement on this point.

In F&C Schlick still takes knowledge to be the reduction of one thing to another, in much the same way as he did in AE (F&C, pp.314-317), but this fact isn't part of the argument for his position. He doesn't actually bring up knowledge at all until the second lecture. We saw already (section 2.2.2) that the interconnection of judgements was originally thought to be a necessary feature of knowledge rather than meaning, and that in F&C it had become part of what it is for a sign to be a statement. More important now than knowledge is expressibility, and the central thesis of F&C - that the content of experience is inexpressible - doesn't require him to bring in knowledge at all. In AE Schlick arrives at the view that concepts are defined by their interrelation by tracing backwards from how knowledge involves recognising objects as falling under concepts, through the way concepts are defined, to the point of primitives, like the concept of *blue*, where definition fails. In F&C, on the other hand, he starts from the view that expression is possible and works forward to the necessity of an underlying structure for propositions. As a result, the most obvious difference between Form and Concept is the way concepts were presented as dependent on definitions, while logical form can be something which is never explicitly stated.

A related point is that Schlick in F&C no longer makes the distinction between everyday language and the language of science - something which shows up

nowhere in the *Tractatus*. People don't use mental images as a rough approximation for concepts. There is no need for the rigorous definitions in the background when the logical form is shown forth by grammatical rules of use with which all competent language-users are familiar.

Ultimately it seems like Form and Content map very neatly onto the ideas of Concept and Intuition which were present in Schlick's earlier philosophy. The key difference is in the way the importance of these ideas is fleshed out. Rather than approaching them through epistemology, Schlick now makes use of Wittgenstein's picture theory of language to explain how logical form is a necessary underlying feature of expression and why content can never enter into it. The epistemological ideas of AE are retained, but their significance is in explaining why the incommunicability of content is no great loss. The idea that logical form, and not intuitive content, is the source of meaning for judgements/propositions was present in AE, but it is much stronger following Schlick's linguistic turn. Where before, intuition was merely too imprecise for science, now it plays no role in meaning determination for either the scientific or the everyday language.

2.2.6 Communication of Form

We have said that expression consists in matching the form of the fact in another fact - the statement - which is in turn matched in thought by a reader who understands the grammar of the language in question. If all that is transferred from one person to another is logical form, does this mean that there is no difference between understanding, for example, "the electron moved from *A* to *B*" and understanding "the electron moved from *B* to *A*"? Both sentences have the same structure, and it is at least plausible to think that two spatial points relate to each other in roughly the same way (i.e. each is located the same distance away from the other). What Schlick's claim seems to amount to is that we could translate either as "$\exists x\ \exists y\ \exists z\ (Rxyz)$" where *ReAB* is to be interpreted as the relation between an electron and the points it is moving from and to – this is the result of replacing "the electron", "*A*", and "*B*" with metalinguistic variables about which we know only that they bear some relation to one another. The problem with this is that it appears as

though if we knew that $\exists x \exists y \exists z$ (*Rxyz*) and were then told "and the electron emitted a photon at *B*," then it's not clear how we could relate the new information with the old. Conjoining our information together, we now know something like $\exists x \exists y \exists z$ (*Rxyz*) & $\exists x \exists y$ (*Sxy*), but that is clearly less informative that the natural-language sentences we started with. The variables are bound by different quantifiers, so we can't say from the formalised information whether the photon was emitted at the electron's start point or end point. This implies that it is impossible to adequately incorporate new information into our network of beliefs, which is not good for a theory of science.

If this is indeed a problem, then it would appear also to arise in the TLP:

> One can describe the world completely by completely generalised propositions, *i.e.* without from the outset coordinating any name with a definite object.
>
> In order then to arrive at the customary way of expression we need simply say after an expression 'there is one and only one *x*, which …' and this *x* is *a*.
> - TLP: 5.526

Here Wittgenstein is saying even more explicitly than Schlick that we should be able to express everything as purely logical relations. Nothing is lost by removing the specifics.

There is, however, an important difference between Wittgenstein's statement and the way I have portrayed Schlick's position. Wittgenstein is specifically saying that completely generalised propositions are adequate for a *complete* world-description – it doesn't follow that they are adequate for single propositions on their own. A complete world-description precludes the possibility of bringing in new information. The problem which arose for Schlick's view here was that, once we've taken away the specifics of which variable was filled in for which position, it becomes impossible to interpret new information. For Wittgenstein, however, there can be no new information - we know that the photon is emitted at

B before we generalise the proposition "the electron moved from *A* to *B*", so we know that the photon is emitted at the electron's final position. That is, instead of having $\exists x \exists y \exists z\,(Rxyz)$ & $\exists x \exists y\,(Sxy)$, we have $\exists x \exists y \exists z\,(Rxyz \,\&\, Sxz)$, so some of our variables are bound by the same quantifier.

The problem is closely related to Carnap's objection to early-Schlick's use of implicit definitions, and to his use of structure descriptions in their place. The key to making structure descriptions work is that the entire network is specified at once. Hence, there can be no new information that we need to incorporate into the system. Carnap's position was that it is possible to give an extensionally adequate structural schematisation of science, and this does not require that we should be able to know in advance how some additional structural points or relations would be incorporated into the system. The problem Carnap identified with implicit definitions was that they defined an empty structure which couldn't be related to the everyday language of experience without some additional stipulations. After taking this criticism on board and being exposed to the ideas of the TLP, Schlick moved to a view on which concepts are defined by their position within a complete network of formal relations beyond which no information could possibly exist. For Wittgenstein and later Schlick, understanding the meaning of a term consists in understanding its internal properties, as we will elaborate further in the next section. Internal properties are those that necessarily relate the term to other terms, and that means that understanding a proposition involves understanding how it relates to the rest of the propositions in our network of beliefs, including those about immediate experience. What is *not* involved in the understanding of a term is any kind of intuitive content. In the example above, understanding how to relate "the electron emitted a photon at *B*" to the knowledge that $\exists x \exists y \exists z\,(Rxyz)$ doesn't require knowing which of y or z can be found at some objective Newtonian coordinates that truly are *A* or *B*. Instead, it requires understanding which of y and z bear the same relations to other terms in our network as the point where the photon was emitted. In this way, it becomes possible to link all of the terms in our language to each other, including experience terms, without at any point relying on the intuitive phenomenal character with which we might associate those terms.

Wittgenstein's conception of understanding as a knowledge of formal relations is what helps Schlick avoid the objection that we saw Carnap raise at the end of chapter 1.

2.3 Moving on from Implicit Definitions

In chapter 1 we examined Schlick's use of implicit definitions and saw that they would require some modification along the lines that Carnap was to suggest in the Aufbau. As we noted, Schlick abruptly stopped talking about them in 1927, despite having until then considered them "absolutely fundamental and of immeasurable import for the whole of philosophy," (1926, [1979b, p.102]). In this section we will examine how Wittgenstein's account of the internal properties of terms allowed Schlick to reintroduce to science ostensive definitions which he had previously rejected.

Both Schlick and Wittgenstein agree that the meaning of most of our everyday words is settled by explicit definition in terms of other words. The question is: what happens when we run out of other words? How do we define the most simple words for the most simple objects? For the sake of example, I'm going to pretend that colours like red are objects.

Wittgenstein's view in the TLP, which we examined above on pages 34-35, was that the proposition "this is A" would serve to explain the meaning of "A" to someone. Aside from the imprecision, this introduced a problem which Wittgenstein was aware of by the time of his *Philosophical Remarks* (1930, [1975]), and which Schlick had written of in 1925 (AE: pp.75-76). The problem is that the same utterance *cannot* be a proposition and a definition at the same time. A definition is a rule. It is not true or false, it merely sets up an arbitrary relation between symbols (or between concepts for early Schlick). On the other hand, it is an essential feature of propositions that they can be either true or false. The difference becomes clear when we analyse the sentences in question, as we did in chapter 1, pages 17-18. "All

bodies are heavy" will be equivalent to "all extended heavy things are heavy" in the case of a definition, or "all extended things are heavy" if it's a proposition. If "A" gets its meaning from the sentence "this is A" then that sentence could be analysed as "this is this". On the other hand, if "this is A" is a proposition then we must take "A" to mean something in particular, which *this* may or may not be.

Wittgenstein's view in the *Philosophical Remarks* is that "in a certain sense, the use of language is something that cannot be taught, i.e. I cannot use language to teach it … And that of course is just another way of saying: I cannot use language to get outside language," (1930, [1975, §6]). The idea here is that the attempt to explain how we learn language by reference to propositions that would teach us was misguided from the beginning. The manner in which we learn the names of objects cannot be through grasping a proposition. It must be outside of language, and that puts it beyond the scope of a linguistic inquiry. Within the language, words for complexes are defined in terms of names for the objects which make up said complex, and that is all that we need to say about that.

Meanwhile, Schlick is moving from a position on which some terms are defined by their intuitive content and some by their position in the axioms of science, to a position on which *all* terms are defined by their relations to other terms in an interconnected network. The big difference between Schlick's earlier and later positions, as far as definitions go, is that in AE he still accepts that some concepts will be defined concretely. For early Schlick, there is an everyday language of experience which we must coordinate with the precise scientific concepts but which itself is imprecise and insufficient for scientific purposes. By F&C, however, formal relations form the basis of definition for not only the abstract concepts of science, but also those terms that appear to be defined by the intuitive experience with which they are associated. We will be looking at structural definitions for terms like "green" and "warm" alongside less controversial theoretical terms such as "photon" and "energy".

For Wittgenstein, names denote objects and resist definition of any sort. The first point to note is that there is nothing imprecise about the denotation of a name

being an object. Schlick's objection to concrete definitions in science was aimed at the fixing of the reference of a name by a recollected sense image, which is not precise, but vagueness is not a feature of reality. The question is how we come to *set up* the relation between the name and the object, and in F&C Schlick makes it clear that he thinks we learn language by having the situations in which propositions are verified indicated to us. That means presenting us with the objects named in the relevant proposition.

The important idea introduced by Wittgenstein is that of an object's internal properties. Objects, by their very nature, show something about the possibilities of their combination with other objects – this is what is reflected in the grammar of a word, which tells us the possibilities for combining the word with other words to form a meaningful sentence. What this means is that it is possible to learn something meaningful about objects by being presented with them, without claiming that the intuitive content of the experience of the object is part of the meaning of the word thereby learned. Being presented with a green leaf and being told "this is green" allows us, after a few more ostensive acts clarify that it is the colour and not the shape that we're interested in, to attach a meaning to the word "green" based on the internal properties of green, not merely on the phenomenal content of greenness.

The difference between this procedure and the procedure of AE (and the procedure of structure descriptions suggested by Carnap) is that we can understand "green" as referring specifically to an object with such-and-such internal formal properties, and not merely as whatever in fact fits into a certain place in the system. The system is important for knowledge, but the meaning of a word is determined by the internal properties of the object which that word names, and these properties can be learned in the process of ostensive definition.

2.4 Conclusion

As we move into the analysis of F&C which will be the subject of the next chapter, here is what we already know about Schlick's position.

First and foremost, knowledge is articulate. The kind of thing which can be known has two parts – the thing that is known and the thing that it is known *as*. Intuitive content is not articulate – you don't learn anything *about* greenness by experiencing it – so it is not the kind of thing which can be known or expressed. The subject matter of knowledge and language is the network of relations which objects bear to one another.

The truth of a proposition consists in the unique coordination of that proposition with a matter of fact. A structure description of a sufficiently variegated world (which we may assume our world is) will be able to achieve this unique coordination because there will only be one way to assign actual objects to the nodes of the structure in such a way that the structure of the world matches the description. Hence an entirely formal language comes with truth conditions which relate to the real world.

An individual proposition (as opposed to a network) says of the world that the objects named in the proposition are arranged in the way that they are arranged in the proposition. That is, a proposition in some sense matches the structure of the world that it talks about. The meanings of individual words are the objects they denote, and their grammar matches the internal properties of those objects so that they can be combined into propositions in the same way that objects are combined into facts. The meanings of individual words can be learned by ostensive definitions, which allow us to learn the grammar of the words because the internal properties of the objects are thereby shown to us. This process does not, however, forge any link between the intuitive content of the experience of the ostensive definition and the meaning of the word learned.

As we explore Schlick's position in F&C it will be clear that it is closely related to the view which he laid out in AE, but also that the ideas of Carnap and Wittgenstein have been taken onboard.

3. Form, Content, and Grammar

3.0 Introduction

In this chapter I will explore Schlick's notion of Grammar and his post-Wittgenstein philosophy. I'll have a particular focus on his 1932 "Form and Content" lecture series, but for questions of interpretation which are not settled by the transcripts alone I will be treating Schlick's later philosophy as sufficiently unified that other papers can provide answers. We will see how, in addition to the syntactic rules of language which were accepted by Carnap and Neurath, Schlick thought that languages required semantic rules of use in order for someone to learn the language in the first place. In the second section I will lay out the details of Schlick's Form & Content philosophy and we will see how it combines with the aforementioned semantic rules to produce a picture theory of language. In the third section I will consider the objections which Schlick faced from his contemporaries and the arguments given by Waismann and Oberdan that Schlick must have moved on from many of the views expressed in F&C by 1934. I will argue that these views were and continue to be based on a misreading of "Form and Content" as an attempt to express the ineffable, when in fact Schlick tried no such thing.

3.1 On Schlick's Conception of Grammar

It is impossible to get a good grasp on Schlick's notions of Form and Content without introducing his wider views on how rules of use entered into language. In this section I will explain Schlick's conception of grammar and why, for Schlick, all languages had to be grounded in empirical data.

Schlick made it clear from an early stage that his conception of language departed from Carnap's syntacticism. Carnap, at the time of the protocol sentences debate, argued that all of the relevant features of a sentence could be found in its

syntax. On this view the truth of a statement was determined by the relation that statement held to other statements (and particularly to the epistemically privileged protocol sentences, at which we will look more closely later), and not by the relation that statement held to the world. For Schlick this was unacceptable - he had held a correspondence view of truth since at least 1910 and viewed rules of use as an essential feature of languages. Even after Carnap (with a considerable push from Tarski) endorsed a semantic approach to language he still thought you *could* construct a language without rules of use (albeit not a particularly useful one) (Oberdan, 1996, pp.288-289; Carnap, 1950, [1956]), but Schlick argued that it was impossible to construct such a language. Though the rules of grammar are conventionally chosen, Schlick argued, you must choose *some* rules, or you have failed to create a language.

Grammar, for Schlick, is a system of rules for how meaningful sentences are to be constructed in a particular language. It is important to Schlick that the system not only includes rules for how words relate to each other, but also includes rules of use which relate words to reality. Waismann gives his own explanation of grammar as follows:

> By 'grammar' we mean everything to do with language that can be laid down even before language is employed. The opposite to it will then be the employment of language. If we take the word 'grammar' in this broad sense, it will include not only the ordinary rules of linguistics, … but also, for example, the definition of simultaneity, the rules of calculation in mathematics and of inference in logic, and, besides these, indicative explications of the form 'this colour is called yellow', where we point to an instance of this colour … in short, grammar will include all the immense number of arrangements presupposed, though nowhere expressly formulated, in the understanding of everyday language.
> - Waismann, 1938, [1979b, p.xxvi]

Schlick himself goes into detail in "Meaning and Verification" in which he confirms that grammatical rules consist of definitions, some of which relate words to each other (i.e. ordinary definitions which constitute syntactic rules) and some of which relate words to their use (i.e. ostensive definitions which constitute semantic rules).

> The 'grammatical' rules will partly consist of ordinary definitions, i.e., explanations of words by means of other words, partly of what are called 'ostensive' definitions, i.e., explanations by means of a procedure which puts the words to actual use. … It is clear that in order to understand a verbal definition we must know the signification of the explaining words beforehand, and that the only explanation which can work without any previous knowledge is the ostensive definition,
> - Schlick, 1936, [1979b, p.458]

We will tackle the two kinds of rule in turn.

3.1.1 Syntactic Rules

Syntactic rules are the rules which govern how words relate to each other. For example, there are rules which tell us how a sentence can be constructed. Syntactic rules tell us that predicates like "blue" must be combined with a noun phrase like "the ball" to constitute a grammatical sentence, and they tell us that the sentence fragments into which "blue" can be combined will be the same sentence fragments as can accommodate “red” or “green”. It makes no sense to say "the blue" on its own, or "the blue was loud", or "the loud was blue", but it makes sense to say "the object was blue and loud" or "the blue of object *x* was darker than the green of object *y*", and it is syntactical rules which tell us which kinds of things can be combined with each other in this way.

There are also syntactic rules which tell us how sentences relate to each other, such as "if S1 has the form '*p*' and S2 has the form '*p* implies *q*' then S1 and S2 let us infer S3: '*q*'". These are the rules of inference in logic and in mathematical systems.

Another kind of syntactic rule is composed of rules like "if *x* is uniformly coloured and *x* is red then *x* is not green". This kind of rule tells us that the sentence "*x* is uniformly-coloured and *x* is red and *x* is green" is nonsense despite being a well-constructed sentence in the sense that there are no obvious failings in the surface structure. (Rules of logic are not sufficient for this purpose because not all mutually exclusive predicates exclude each other logically, see for example Wittgenstein's failure to get colours to exclude each other (Wittgenstein, 1929; Wittgenstein, 1931, [1969, pp.210-211], referenced in Uebel, 2007, pp.85-86).)

Although we might be inclined to think of them as semantic, for our purposes definitions also fall under the category of syntactic rules. "A bachelor is an unmarried man" is a rule which tells us things about the interpretation of sentences containing the word "bachelor", even if there are no unmarried men in the world, and it makes no sense to say "he is a bachelor and he is married". This is a fact about the relation of words to each other; not words to the world as Schlick's semantic rules are.

The unifying feature of syntactic rules is that violating them leads immediately to nonsense. You cannot form a meaningful sentence which combines the word "blue" with a kind of thing that colours cannot be applied to; you cannot form a meaningful sentence which violates metaphysical necessity; you cannot form a meaningful sentence which asserts something false by definition; and you cannot meaningfully assert two mutually contradictory sentences. There are certainly more kinds of syntactic rule than I have given examples of here, but they will all be recognisable by the way that sentences which violate them make no sense.

An interesting feature of syntactic rules is that Schlick believes it is nonsensical to assert them. This is because Schlick believes that for it to be meaningful to assert *p* it must also be meaningful to assert not-*p* (F&C, p.300). Rather, syntactical rules show themselves in what kinds of sentence it is possible to meaningfully form. This position is closely related to Schlick's endorsement of Wittgenstein's view that we cannot get outside of language to talk about it (see pages 50-51).

3.1.2 Rules of Use

Schlick's first clear statement of the importance of rules of use comes at the end of the first "Form & Content" lecture. There he argues that the only satisfactory answer to the question "what do you mean by *p*?" will be in some sense a replication of *p*. "An explanation which said *more* than the proposition itself would not be a correct explanation of it, and if it said the same thing as the proposition [i.e. was merely a repetition of *p*] it would be superfluous," (F&C, p.310). Schlick's point here is that when we explain the meaning of a sentence, we cannot say anything except what the original sentence said in different words. An explanation that failed to use different words would obviously not be helpful - the entire reason we're explaining the meaning of the sentence is that our interlocutor didn't understand the sentence as it was originally phrased. And we couldn't give as explanation any sentence which didn't mean the same thing as the original sentence, otherwise we would have defined the meaning of the sentence as something which was not the meaning of the sentence, and down that road lies madness.

So, when a proposition is expressed in words unknown by the interlocutor, it must be rephrased in language he can understand. If he does not know a language, or the language he does know is inadequate, then the only way to teach him will be to indicate the state which the world must be in to make the proposition true. This, necessarily, means that language will be linked to a state of affairs - if it wasn't then it could never be taught, and clearly it is teachable since we do in fact teach it.

> [T]here is not the slightest mystery about the process by which a sentence is given meaning or turned into a proposition: it consists in defining the use of the symbols which occur in the sentence. And this is always done by indicating the exact circumstances in which the words, according to the rules of the particular language, should be used.
> - F&C, p.310

That is to say, rules of use are the difference between a sentence being a mere string of symbols and that sentence being used to express something (we will go into further detail about the nature of expression in section 3.2).

In contrast to syntactic rules, violating a semantic rule does not lead to nonsense. Forming the sentence "this is red" when faced with something blue is entirely meaningful. It's just false. Violating grammar always leads to a defective sentence, but the kind of defect you end up with depends on the kind of rule you violated. (Violating rules from both sets at the same time leads to nonsense, so it could be argued that in a certain sense the syntactic rules take precedence. Sentences must pass the syntactic test before they get exposed to the semantic one.)

Here we might ask, what became of the structuralism that we have been encountering since chapter 1? We have seen how Schlick tried to define the concepts of science in such a way that they would not depend upon a relation to the world. Now it looks like he's going back to an account on which meaning is determined by experience and linked in some way to intuitive content.

As we'll soon see, Schlick certainly does not renounce structuralism at this time. We have seen that in AE Schlick attempted to ground meaning in implicit definitions, but these turned out not to be sufficient to fix the properties of concepts. We then saw that Carnap offered structure descriptions as an alternative, which would fully determine the state of the world by specifying a full network of objects and relations. This showed that it was possible to describe the world in fully structural terms, but Schlick came to see that the *Aufbau*'s method of explicitly reducing concepts to a single basic relation would be problematic if Wittgenstein's account of language was to be more-or-less adopted (Uebel, 2007, pp.84-90). We saw in the last chapter that knowing the internal relations of an object is what knowing an object consists in (see pages 31-32; also F&C, pp.294-295), and we've said that the syntactic rules for a term will correspond to the internal relations of the object, however we've also seen that Wittgenstein objected to the idea that we can get "outside of language" in order to define its limits. The upshot of this is that

we cannot actually state the syntactic rules that constitute the meaning of words.[16] Therefore, to learn a language in the first place, a subject needs to be *shown* rather than told something. Showing the subject an object is a way of conveying that object's internal properties to them, and connecting it with a word thereby teaches the syntax of that word. The semantic and syntactic rules therefore work together. Semantic rules are teachable through ostensive definitions (which play the role of the elucidations we saw in Wittgenstein's system), and it is from our understanding of the semantic rules that we are able to grasp the syntactic rules.

3.1.3 Phenomenalist or Physicalist Rules of Use?

There are a couple of ways in which we might interpret these rules of use. The physicalist approach would be to say that the rule *'x is yellow' is true iff x is yellow* means that it is true to assert "*x* is yellow" about all and only those objects which are in fact yellow. On this interpretation "daffodils are yellow" is true even when we look at the daffodils through a green filter. The phenomenalist interpretation, on the other hand, would have it that *'x is yellow' is true iff x is yellow* means that it is true to assert "*x* is yellow" about all and only those objects which appear yellow. To be yellow is to be perceived as yellow. In other words, "yellow" refers to the subjective experience rather than the objective wavelength of light reflecting from a surface.

Given that Schlick's argument for rules of use rests on the fact that the language is learned by experience, we might be inclined to think that the second interpretation is more likely. If we learn how to use the word "yellow" by being shown yellow things then it seems like what we will associate with "yellow" is the way those things appear, and not whatever intrinsic property they have which makes them appear that way. But this would introduce problems for talking about

[16] Nowadays, and throughout this thesis, we have no problem drawing a distinction between object language and metalanguage and defining rules for the object language in the metalanguage, but for the Vienna Circle it would be a few more years before this way of doing things was widely adopted.

the world. For one thing, when we talk about the colour of something, we don't want to just talk about how it appears right now - even when we're looking at them through a green filter, it seems we want to say that the daffodils themselves really are yellow - we want to talk about what colour it actually *is*. There is no obvious contradiction in the sentence "this is yellow, but it appears red". Another problem arises with the possibility of intersubjective communication about subjective experiences. If "yellow" refers to my experience of yellow, it is not clear that anyone else will understand what I mean when I use it.

These are genuine problems for a grammar founded on phenomenalism, but it would be a mistake to interpret Schlick in this way. Schlick, at the time of the "Form and Content" lectures, was arguing that phenomenalist and physicalist approaches to understanding the world are qualitatively indistinguishable from each other and so it is meaningless to draw a distinction between them. This does not necessarily mean that something is yellow if it appears to be yellow - for Schlick the meaning of a statement is a potentially infinite set of conditionals of the form "if these particular circumstances obtain, then this particular experience will occur" (1932a, [1979b, p.280]). This means that appearing yellow under a yellow filter is not sufficient to make something yellow - it must also appear yellow under white light with no filters and no other adverse conditions. The fact that an observation is corroborated by other related observations is what separates illusion from reality (F&C, p.273). Schlick's interpretation of sentences is problematic because under classical logic conditional statements are true so long as the antecedent is false, meaning that "if x is observed in ideal conditions, x will appear yellow" is true so long as we don't in fact observe x in ideal conditions. But this problem need not concern us here because, regardless of the problems which arise for Schlick's particular brand of verificationism, what we're looking for at the moment is the correct interpretation of Schlick's notion of grammar, and his attempt to break down the physical/phenomenal distinction gives us that. It seems that Schlick understood rules of use as telling us how to correctly use terms based on both how the world really is *and* how it appears to be, because, under sufficiently close inspection, there is never any question of the world appearing to be some way that it isn't. That is to

say, if there is no evidence to be found that something isn't yellow, then it really *is* yellow.

3.2 Form and Content

Having established that rules of grammar including both semantic and syntactic rules were an important part of Schlick's philosophy, we will now move on to look at Schlick's philosophy of Form & Content. It will be obvious as we go through that Schlick has to a great extent adopted Wittgenstein's picture theory of language.

3.2.1 Expression and Representation

The first crucial point in Schlick's argument is the distinction between Expression and mere Representation.

Expression - a relation between the fact that a certain utterance (an instance of something that falls under the category of Language) has been made and that possible fact which the utterance says obtains. For example, some black marks on a piece of paper are able to teach someone that a volcano erupted on a faraway island, even if the reader has never seen that particular arrangement of black marks or the eruption before. The relation between the black marks and the eruption is Expression - the black marks express the possible state of affairs that constitute eruption (whether or not that state of affairs actually obtains). (Of course, lots of relations will hold between the black marks and the eruption - the relation of the two facts being a certain distance apart, for example. Schlick presumably means Expression to be the relation which necessarily holds between all facts that express and the respective possible facts that they express. Shortly, we will be exploring the exact nature of this relation, but we cannot elucidate it at this stage because *that expression works like this* is the very thing we're arguing for. For now, we can only assert that there is such a relation.)

Representation - an arbitrary correspondence between one thing and another, whereby we agree the one shall stand for the other in certain contexts. For

example, by convention the various lines on a map represent roads and rivers and so on.

Words, according to Schlick, are things which, by convention, represent objects in the world: "our words and all our signs for words are symbols which, partly by arbitrary agreement and partly by accidental usage, stand for the things of which they are symbols," (F&C, p.287). Sentences, on the other hand, must express, rather than represent, because otherwise we wouldn't be able to recombine the elements of sentences we have heard before into new sentences which express new facts. Speech is not just "the simple repetition of signs whose meanings have been learned by heart," (F&C, p.288) because grammar, as well as vocabulary, is part of any language, and if you only learned that certain sentences with certain constructions represented certain facts this would not be enough for you to take apart those sentences and recombine them to create new sentences. For example, I can know that Δ represents a particular sound without having any idea what ∇ (i.e. the same symbol rearranged; in this case, turned upside down) would mean. But if I know what "the book is on the table" expresses then I can't help but know the meaning of "the table is on the book". Of course, if "the book is on the table" were to be turned upside down I would have no idea how to interpret it. In this case the representing symbols - the words themselves - would have been rearranged, and rearranging a representing symbol usually results in nonsense.

It would also be possible to rearrange "the book is on the table" in such a way as to render it unintelligible without rearranging any individual word, for example as "table the book is on the". The reason for this is that it's not quite true to say that *sentences* express facts. That which expresses a fact is a *statement*. A statement is a sentence coupled with the rules of grammar which are needed to interpret it. In "table the book is on the" the syntactical rules used to interpret the original sentence have been broken, so the words cannot be interpreted and the sentence fails to form a statement. So, an expression cannot be rearranged in just *any* way to form a new expression. Some expressions may resist any rearrangement at all. For example, "John is a man" obviously expresses a fact, but it cannot be rearranged to express a

different fact without at least the addition of some punctuation. So, when Schlick says that "the only essential point" distinguishing Expression from mere Representation is that…

> The same set of signs which was used to describe a certain state-of-affairs can, by means of a rearrangement, be used to describe an entirely different state-of-affairs in such a way that we know the meaning of the new combination without having it explained to us,
> - F&C, p.289

… he isn't saying that a meaningful sequence of symbols is an expression only if it can be rearranged to form some other meaningful string of symbols. He is making a much more general and uncontroversial claim that expressions can be taken apart and recombined with bits of each other to make new, significant expressions.

Are expression and representation exhaustive?

That is to say, does *everything* fall under either the category of representation or that of expression? The answer to this is evidently "no". For a thing to represent something else a convention must be introduced, and there are obviously many things in the world which we have not decided will represent anything. For example, pebbles, blades of grass, etc.. Until someone decides that they are to represent something, they will not.

A more interesting question would be whether or not everything *which is used in language* either represents or expresses. "Language" "denotes *any* system of things or procedures or events considered as a means of communication of thoughts" (F&C, p.286, original emphasis). The system of written English seems to include not just words and sentences, but also individual letters. Does "y" fall under either category of expression or representation?

We might want to argue that there are actually two systems in play when we look at written English. On the one hand we have English, in which words represent and certain complexes of words - sentences - express. On the other hand we have

the Latin alphabet, in which letters represent bits of sounds, and complexes of letters - words - express pronounceable syllables. But sounds can be rearranged in just the same way that letters can, so that just shifts the question down the line (or round the circle, since we could just as well suggest that sounds represent letters).

Whilst it is true that letters can be split up and rearranged to form new words, words on their own are not significant utterances. We cannot recombine letters and immediately understand the new combination purely by knowing the rules of grammar, because the rules of grammar do not tell us anything about how letters may be arranged. As such it seems fair to say that letters are not part of a system for communicating thoughts until they are arranged into words. If letters on their own were considered part of language then the same reasoning would lead us to say that the black marks which make up letters are part of language too, since they are part of what makes up the elements of a system and they can be rearranged into a different part. But clearly a black mark by itself is not part of a system of language. Black marks have surely existed much longer than any concept of language (e.g. charred bits of prehistoric lightning-struck trees) and an element of language cannot predate the existence of the language itself. Even if a random arrangement of fallen Jurassic-period leaves were to accidentally spell out "we are leaves" we would not consider an element of a language to have predated that language because it would not truly be a statement. The signs can't represent anything, and so can't be words, until the conventional correspondence exists. It might be argued that letters are essential for language where little black marks are not - writing can be done in other colours too. But marks of *some* kind are necessary for writing, and examples of hieroglyphics and various East-Asian languages demonstrate that letters are not a necessary aspect of language.

So, it seems like systems of language are made up of representing parts and arrangements of those parts which form expressing compounds. Anything which doesn't represent or express cannot really be said to be part of a system of language, and as such expression and representation are exhaustive.

Are representation and expression exclusive?

Are representation and expression such that if something expresses it does not represent, and if something represents it does not express? There is a sense in which we might want to say that all expressions represent the facts which they express - a sentence makes us think of a state of affairs in much the same way as a single word makes us think of the thing which that word represents. But if we take all expressions to be representations then, given our previous conclusion that all elements of language fall under the categories of expression or representation, all elements of language will fall under representation. This would make the term "representation" redundant. It would be more useful to think of the term as only applying to symbols which stand in for things, not facts. So, the question is: can a sentence which expresses a fact also at the same time represent a thing?

Consider the following arrangement of sentences:

(P1)	(P2)
This sentence represents object A.	This sentence represents object B.

P1 and P2 are sentences which express the fact that P1 and P2 represent objects A and B respectively. They could be taken together to describe the state of affairs in which A is to the left of B. Does this show that a sentence can be both an expression and a representation?

Once again, more than one system is needed to get this result. The rules of English alone do not use sentences to represent anything, and the arrangement of the sentences above does not express the fact that A is to the left of B unless we introduce rules for interpreting it. We could just as well interpret the arrangement as expressing that A is to the west of B, or A was invented before B, or A is a darker shade of red than B if we chose appropriate rules for interpreting the arrangement. P1 and P2 are expressions in English but only represent in a different system - a system which happens to use English sentences as its representing units and uses

spatial arrangements of those sentences to express facts. So, this example does not show that sentences which express can also represent.

Another way in which we might think expressions can represent is with definite descriptions. Terms like "the queen of England" are composed entirely of symbols which could be rearranged[17] but seem to represent a thing. This does serve to highlight the point that things can be denoted by a chain of words, and that therefore the fact expressed won't always have the same number of parts as the statement has words, but this is not actually an example of a representing expression. This is because the link between "the queen of England" and the queen of England is not one of arbitrary learned correspondence. It's a relation which is set up by the meaning of the words involved, each of which will themselves represent something else (i.e. "England" seems to represent England and "the queen of" seems to be part of the representation of the relation *x is the queen of y*; (see pages 35-36 above and section 3.3.1 below)).

Since Representation is a relation between facts and things while Expression is a relation between facts and facts, for the same fact to both represent and express would be for that fact to have two different interpretations. This might not seem like a problem, because several words in English have more than one meaning. For example, "seal" can refer to a marine mammal or to the act of sticking closed an envelope. We could imagine some eccentric naming their dog "roses are red", in which case it would seem like there was one use of the sentence which referred to the dog (and therefore represented the dog when used in a sentence) and another which described the colour of roses. However, I have been oversimplifying matters by talking about English as a system. Schlick makes it clear in his later work that the rules of language are not supposed to be a single unified system within which we construct statements that are interpreted according to those rules. Rather, a

[17] The term could not be rearranged by itself - "the England of queen" makes no sense at all - but its components can be taken apart and combined with other words, and we said before that this was all that was required of an expression (p.61).

statement is a sentence coupled with the rules for the particular interpretation of that sentence *on that particular occasion*, so that "the sentence 'it was a bold flight' [is] an altogether different statement, when speaking of the act of a fugitive, from when the reference is to an aviator's feat," (Schlick, 1935b, [1979b, p.408]). The rules which we use to interpret a sentence, without which that sentence neither expresses nor represents anything, remove all ambiguity, so there is no longer any question of the same statement having two interpretations. This means that, once again, while the same sequence of symbols might be used to represent something in one context and to express something in another, it can only do so when we are mixing up two different systems. Within one set of rules a symbol, or complex of symbols, can never both express and represent, so the two relations are exclusive.

So, Representation is a relation which holds between symbols (or complexes of symbols) and things. Expression is a relation which holds between complexes of representing symbols and facts. Facts are made up of a certain arrangement of things, which would constitute different facts if arranged differently. Expressions are made up of certain arrangements of representations which would constitute different expressions if arranged differently. Schlick noted the parallels between expressions and facts, and between representations and things, and turned his attention to arrangement, or "Logical Order", as the important feature of expressions.

3.2.2 Logical Form

From the fact that expressions can be rearranged where mere representations cannot be, Schlick concludes that "the essential feature of expression is Order," (F&C, p.290). As will become clear, Order is a necessary feature of expression, but it is not the only one and nor is expression the only thing which has this feature, so calling it "*the* essential feature" might be going a bit far. Nevertheless, it is *a* feature which distinguishes expression from mere representation.

Internal and External Relations

Order, in this context, is not particularly about temporal or spatial arrangement (although those arrangements can be a kind of order) but is about logical structure. It is something which is found in both propositions and the facts which those propositions express. The logical structure of a fact is everything which is communicable about it - size, shape, colour, etc. are all part of the logical structure of a fact. Schlick follows Wittgenstein (TLP: 4.122) in identifying two different kinds of relation which make up logical structure: internal and external.

Internal relations - relations between things such that if the things exist the relation *must* hold between them. For example, that 2 is greater than 1 and less than 3 is an internal relation. "[I]t is nonsense to suppose that instances might be found in which twelve would not be twice as much as six," (F&C, p.294).

External relations - relations between objects which are entirely contingent on the state of the world. For example, that the chair is to the left of the table is an external relation. If the world were a little different then the chair might be to the right of the table.

It appears that external relations hold between things, while internal relations hold between *properties* of things. Schlick gives the example of a green leaf lying next to an ink blotter which is a similar shade of green. The fact that the two things are near each other is an external relation - of course the two things could exist in different places. The fact that the two things have the same colour is also an external relation. Either the leaf or the ink blotter could have existed in a different colour; the leaf could have been older, or the blotter could have blotted different inks. But the relation of similarity between the two shades of green is internal. There is no way that those shades of green could exist and not be similar, because a world in which the leaf and the ink blotter are not similarly coloured is a world in which one of those shades of green does not exist (or, at least, does not exist within the objects which we are considering).

Facts are composed of external relations. The spatial relation between the leaf and the ink blotter, for example, is a fact, and the sentence "the leaf is next to the

ink blotter" expresses that fact. (In contrast, internal relations are intrinsic to an object, so they don't involve a combination of objects, and therefore do not constitute facts.) This is what Schlick means when he says that sentences express logical structure - *at least* part of what is expressed is in the form of external relations, because facts are made up of external relations so if the sentence does not describe any relations it does not express any fact.

3.2.3 Matching Logical Structure

Schlick's notion of structure might appear to be a little confused at first glance. First he speaks of the sentences which express facts as facts themselves, i.e. as the marks on a piece of paper. Then later he writes that "the same fact may be expressed in a thousand different languages, and the thousand different propositions will all have the same structure, and the fact that they express will have the same structure, too," (F&C, p.290). "Proposition" means the same thing as "statement" for Schlick, so a proposition will be a fact - an arrangement of marks on the paper or uttered sounds. But in this case, we might ask, how can the propositions in the different languages have the same structure as each other, let alone as the fact which they are trying to express? Structure, it seems, includes shape and sound, and two sentences expressing the same fact in radically different languages are not going to have the same shape.

Schlick writes that expression "implies two facts: one that expresses and one that is being expressed. The former is a sort of picture of the latter, it repeats its structure in a different material," (F&C, p.300). The structure of an object is not simply an amalgamation of its size, shape, colour, etc., but rather it is a system of logical relations which can be brought out in a number of different ways. Colour can, according to Schlick (F&C, p.296), be represented spatially as a double cone, with the relation *similar to* being replaced by *close to*. Alternatively, it could be replaced by a system of written words, with the relation *similar to* being replaced by certain grammatical rules for the use of the term "similar to". The sentence "the volcano erupted" does not have the same structure as the eruption of the volcano in terms of how the two facts look, but rather the structure of how the erupting volcano looks

is to be found in the ways of meaningfully using the words "volcano" and "erupted" in conjunction with visually descriptive words. That is, the logical structure of a sentence depends on what grammatical rules are being used to interpret it, just as what is logically entailed by a set of premises depends on the system of rules of inference you choose. Sentences match the logical form of the facts they express when interpreted according to the rules of the language they are formed in, but not when interpreted according to the rules of geometry.

3.2.4 The Relationship Between Form and Grammar

This suggests another way of thinking about the semantic rules from section 3.1.2. They are not just rules which say "in such-and-such circumstances it is correct to assert so-and-so". They are translation rules which let us convert the logical form of a fact into a new material. The rules of English let us translate the fact that there is a chair next to a table into the series of black marks: "there is a chair next to a table", preserving the logical form of the fact but giving it a new material.

We might be tempted to say that rules of use such as "*this* is called 'green'," (simplifying from something like "*x* is called 'green' iff *x* is sufficiently similar to *this* along the dimension of colour") give us more than just the logical form of "green". It seems to involve some particular indescribable sensation - the kind of thing which we might want to attach the word "content" to. We will say more about Content below, but it is important to see that semantic rules do not depend on it. We've already said (see section 3.1.2) that what a rule of use does is to show us the internal relations – the logical form – of the object named, and thereby allow us to see how the word thus defined must combine with others in the proposition. Schlick emphasises this point by considering a case where we attempt to communicate the colour of a leaf by literally presenting the leaf to the subject. Even in this case, the rule which we are considering will not attach the word "green" to the sensory content of the experience of greenness, because the sensory content of the experience of greenness is not in any way transported along with the leaf. What we have transported is a property of the leaf which relates it to other things that we can visually experience – a property of similarity with grass and contrast with cherries,

or, more formally, a location in the dimension of possible colours. This property is in no way bound up with the sensory content of the experience. It doesn't matter if one person experiences colours in the way another person experiences sounds so long as they both apply "green" to the same objects and understand that if something is green then it isn't red. (We will see later that it is nonsense to talk about two people experiencing the same thing differently, but for now it is useful nonsense which illustrates a point.)

3.2.5 Content

So far, we have said a lot about Form, even going so far as to say that Form is the only thing which can possibly be expressed, but very little about Content. Partly this is down to the aforementioned fact that, if the view we are investigating is correct, it would appear as though there is nothing that can be said about Content. Roughly, when Schlick says "content" he is talking about that aspect of experience which we cannot describe - it is the feature of colour which cannot be communicated to a blind person, for instance. But speaking literally, there is nothing that cannot be communicated to the blind person because there are no facts about Content over and above its form. Facts are logical relations and Content does not have any logical relations. Content is the stuff logical relations are made of. Schlick resists describing Content because his entire point is that there are no facts about Content to be described, but Oberdan, unconstrained by what he calls Schlick's "implausible, counterintuitive, and highly dubious thesis" (1996, p.279), suggests "the 'greenness' of the colour green, [and] the distinctive odor of the smell of wood smoke" as examples (1996, p.278).

Since Schlick's view is that Form is the only thing which can be communicated there are some traditional questions of metaphysics which he regarded as obviously nonsensical. The example which springs most readily to mind is the inverted spectrum hypothesis - the idea that another person might experience black in the way we experience white, green the way we experience red, and so on. If we take it that the only thing which can be communicated is Form then it makes no sense to ask, "is your red the same as my red?", at least not in the way which is

generally intended. If both of the people in question take "red" to refer to the colour which mixes with the colour they call "yellow" to make the colour they call "orange", both take it to be the colour on the outside of the rainbow, and so on for the other internal and external properties of red, then both of them use "red" in the same way and there are no further questions to be asked.

3.2.6 Summary

This summary of the position appeared in the previous chapter, but I repeat it here because it rounds out the section and because the example sentence S will feature in the discussion that follows.

Facts have a certain logical structure, composed of internal and external relations. Internal relations are those relations which hold necessarily, such as those between numbers, whereas external relations are those which hold contingently. These structures will be naturally thought of in terms of the senses by which they are detected, but they can be translated without loss of structure into any other material we wish by applying suitable rules. Words are one such material, which can be combined with the rules of grammar to form language. Sentences in conjunction with rules are able to express facts by matching their logical structure in a new material (the new material is important, because matching the structure of a fact in the original material would not be an expression, or a "picture" of the fact - it would be "the original itself, or perhaps an exact duplicate of it," (F&C, p.300)).

To illustrate the picture, consider:

(S) The sky is blue.

S is made of black marks on paper (or on a screen, depending on your viewing choices) arranged spatially, and it expresses the fact that the sky is blue. S is not made of the same material as the fact which it describes - the sky itself is obviously not an arrangement of black marks on paper. The letters "T h e s k y" together appear to represent the sky, and the letters "i s b l u e" seem to indicate an external relation of similarity between the sky and other blue objects (or a relation of a particular kind of difference between the sky and non-blue objects) as dictated by

the semantic rules of English. The syntactic rules of English tell us that when "is blue" follows "The sky" the writer is expressing the fact that the property of blueness belongs to the object which is the sky. The structure of blueness (i.e. as a property which is possessed by all and only those objects whose colour is between green and purple) is mirrored by the rules for using blue in a sentence (i.e. it must be ascribed to an object and the objects to which it is ascribed cannot be said to possess a colour between red and yellow (or at any other non-blue position on the spectrum)). And an aspect of the structure of the sky (i.e. having an external relation which places it between grass and amethyst stones along a certain dimension) is mirrored by the way the words “the sky” are slotted into the object position of the “is blue” predicate. The fact that the sky fits between grass and amethyst along the colour dimension is what makes S true when interpreted according to the rules of English. S would be false if it was interpreted with the rules of another language where, for example, “is blue” represented the relation of being between red and yellow, so the choice of language is important. Equally important are the facts of the world – if we imagine a possible world which is identical to our own with regard to how amethysts and grass and all other objects except the sky relate to each other with respect to colour, but in which the sky does not fit between amethysts and grass along that dimension, then we are imagining a world in which S is inconsistent with the other reports of colour which are true in this world. (I hesitate to say that S is false in this world, because we might want to say that S is the only colour report which is true in both this world and that world, and every other object has shifted some way along the colour spectrum. Both are perfectly valid ways of describing that scenario.)

3.3 Objections

In this section I will consider some arguments for rejecting Schlick's Form & Content philosophy, including the view that Schlick himself rejected it within two years. I will not go into objections to Schlick's conception of grammar here, because there is

no question of Schlick having dropped his ideas on that subject and so there is no question of reconstructing Schlick's philosophy without it.

3.3.1 The Possibility of Repeating Form in New Materials

One worry we might have with this picture is the fact that it relies on the translatability of structure from its original material to any other. Schlick's view is that points on the colour spectrum could just as well be represented spatially as visually, or indeed in any other manner you chose. Schlick explicitly asserts that it would be possible for someone to experience colours in the same way that we experience sounds (1932a, [1979b, p.270]) (this was in his essay "Positivism and Realism" in which he was speaking very loosely - it is, strictly speaking, nonsense to question whether or not one person is experiencing colours in the same way as another, since neither a positive nor negative answer has any effect on observation). But we might worry that something like taste is too rich an experience to be reduced to a spatial structure, let alone anything we could put into words.

Assuming reality is more or less how it appears to be, this cannot cause a problem for Schlick. Sense data are transmitted from organs to the brain via neurons which either do or do not fire at any given moment. Therefore, all kinds of sense data can have their structure matched in binary, and from there can be translated into whatever material we choose.

Even if any form can be expressed with any content, we might still wonder whether propositions are an example of this process. Keyt (1964) presents a general worry for this kind of picture theory of language in that there appear to be more parts to sentences than there are parts to the fact which the sentences are expressing. The sentence "the book is on the table" appears to have three parts - two names of objects "the book" and "the table", and the relation "is on". But it seems like the fact only has two objects - the book and the table. What does "is on" represent?

Schlick thinks that terms like "on" represent external relations, and that these terms can be taught ostensively in basically the same way as names can. Keyt argues that if a relation is a thing which can be represented then there will be yet

another part of the proposition to explain, specifically whatever relation ensures that "is on" is to the right of "the book" and to the left of "the table". But given our response to Ricketts' worry (see pages 36-37), we claim that this is to misidentify the elements of the proposition. The elements are not "the book", "is on", "the table", and *things being in the order x, y, z*. The elements are "the book", "the table", and the relation *x is to the left of "is on" and y is to the right*. These things correspond to the elements of the fact: the book, the table, and the relation *x is on y* respectively. The relation *x is on y* cannot exist without two objects in the argument places, and so it needs to be represented in the proposition by something which itself is not grammatical without two arguments. Given a reasonable account of how the picture theory works, worries about the possibility of repeating structure should be soluble. If we have enough freedom with rules of interpretation, we shouldn't have a problem encoding the structures we want in sentences.

3.3.2 Loss of Understanding

One might be tempted to claim that something is lost if we cannot express the content of our perceptions. Have I really understood a person's utterance if I have failed to grasp what he had in mind when he spoke? Schlick, however, argues that there is nothing more that your interlocutor can have had in mind than the logical relations which you have grasped. What other information could there be? A blind man could still be taught all of the facts of colour. If he had a device which could detect colour and then read out loud, for example, a series of numbers corresponding to that colour's place on the spectrum, (and if he had been taught how to interpret those numbers,) then he would be able to learn and express every fact about colour in his immediate environment. There doesn't actually seem to be anything lost unless by "colour" you specifically mean something which is sensed by the eyes, but, as Schlick points out (F&C, pp.298-297), statements about people's eyes and how they react to certain stimuli are also matters of logical relations.

It must be admitted that on this point a lot of philosophers have diverging intuitions. Frank Jackson's example of Mary (1982) is often taken to evoke the intuition that there is something about colour which can only be learned by

experiencing it. Mary is a brilliant neuroscientist who knows every physical fact relating to colour and the perception thereof, but she has been raised entirely in a black and white room and has therefore never experienced colour. One day she is allowed out of the room to experience colours for the first time. Does she learn something new?

There are two points worth making about the example of Mary before we move on. The first is that this is exactly the kind of thing which we saw Schlick arguing against at the very beginning of chapter 1, without relying on philosophy of language at that stage. There is something inherently different between experience and factual knowledge, in that one involves relating two or more things together and the other does not. Given the meaning that Schlick attaches to the word “knowledge” there is nothing that Mary can come to know from the mere experience of a colour. If someone wants to use the word “knowledge” in a broader sense, then that’s fine by Schlick as long as they acknowledge that they’re using the word differently. There is no genuine disagreement here; it’s merely verbal. The second point is that even if you think Mary does learn something, then you still have to acknowledge that whatever it was must have been incommunicable. Mary had access to every fact about colour which could be communicated, so *ex hypothesi* if there was something that she didn’t know then that something cannot be expressed in language. Therefore, the case of Mary is not an objection to the Form and Content account of communication.

3.3.3 Metaphysics

The more serious objection to Schlick's conception of Form & Content is that it seems dangerously close to metaphysics. If the content of experience is ineffable, and if in fact it would be better if the word "content" never entered into our language (F&C, pp.306-307), then it would seem to be nonsensical to say "there is such a thing as content" in the first place. Even Schlick's constant insistences that he is aware of the danger and believes he has avoided it will not save him from contradicting his own words: "the inexpressible cannot be expressed, not even by the philosopher," (F&C, p.291). Ayer objects along these lines - "To maintain that content is

inexpressible is to behave like Ramsey's child. ''Say breakfast.' 'Can't.' 'What can't you say?' 'Can't say breakfast.'' (Ramsey, 1931, [1959, p.325])" (Ayer, 1936, [1959, p.241]).

One possible response to this line of objection would be that the sentence "there is content" involves a misuse of the word "is". "Is", we might think, says that something with the characteristics of Content has a position in the universe. But Content has no characteristics - there are no facts about it - and it never claimed to have a position in the universe. Schlick would be the first to admit that "there is content" is nonsense, and that's why he never said that without qualification. But this response surely doesn't deflect the problem. Admitting that it's nonsensical to say "there is content" is tantamount to admitting that the whole Form/Content distinction was nonsense all along.

The correct response to this line of objection, I think, is to focus more on what expression is. There is a distinction between trying to talk about something and trying to express something. We can talk about objects – describe their properties and so on – but we can't express an object because it's not the kind of thing that gets expressed. We express facts – combinations of objects. Perhaps similarly we can talk about Content despite not being able to express it. Before fleshing out the details of this response, it will be useful to consider whether we *can* do away with Form and Content.

3.3.4 Did Schlick Change his Mind?

Oberdan (1996, pp.279-283) and Waismann (1938, [1979b, p.xxix]) argue that at least by 1935, and perhaps as early as 1933, Schlick himself had abandoned the thesis of incommunicability of content which is so central to the position we have been expositing. Feigl and Blumberg (1974, p.xxi) seem to disagree, taking Schlick's 1925-1936 philosophy as basically unified (and taking Form and Content to be problematically-metaphysical notions throughout that time, but that's neither here nor there). Suppose that Waismann and Oberdan are right, for the sake of argument. It is natural to now ask what we are to make of the arguments which led Schlick to Form and Content in the first place.

Consider again the example of "the sky is blue," = S. In the terms of Schlick's 1932 philosophy, the content of "is blue" is an arrangement of black marks; if we may speak loosely for the sake of argument, the content of the relation sky bears to the visual dimension is ***blueness***. Can we eliminate the notions of Form and Content from our earlier analysis? Presumably we don't want to say that the black marks which make up "is blue" are somehow a part of the logical structure of the sentence. If they were then there would be a difference in the structure of "is blue" and "is blue" (that is, the same words in a different colour), but the two obviously don't represent different relations. If the logical structure of the words depends upon the colour in which the words are written in any meaningful way then "the sky is blue" should express something different to S, but as a matter of fact it doesn't (although of course it *could*, given the right rules for interpreting the language). And we obviously don't want to say that the things which we had previously thought to be content have turned out not to exist at all. The black marks which make up "is blue" obviously do exist. The problem with trying to eliminate Form and Content is that, as Schlick points out, he hasn't really made any claims (F&C, p.300). He's just given names to things which are uncontroversially aspects of language. The only option left for "moving on" from this stage in Schlick's thinking is to say that there are such things as Form and Content, but that the distinction is not the same as that between what can and what cannot be said.

Waismann writes:

> Two later writings, 'On the Relation between Psychological and Physical Concepts' and 'Meaning and Verification', are the first to show how Schlick was able to free his mind from this distinction. In them – again under Wittgenstein's influence – he sketches out peculiar imagined circumstances under which the 'content' of another person's experience would become accessible to us. And with this the distinction between 'communicable' and 'incommunicable' has lost its original meaning.
>
> - Waismann, 1938, [1979b, p.xxix]

It would be worth going through Schlick's arguments in these papers to assess Waismann's claim.

In "The Relationship between Psychological and Physical Concepts", Schlick addresses the charge that some might level against physics that it deals only with the quantitative and not the qualitative, and thereby misses out on something. Physics makes assertions about how things relate to one another without telling us about the relata in themselves – it misses the *content* of its own assertions.

> This content, it is urged, is the psychological. Psychology would therefore confront physics as an autonomous discipline… To the assertion of the one-sidedness and limitations of the methodology of physics, there stands in sharp opposition the claim that an absolutely complete description of the world is possible by the use of physical methods; that every event in the world can be described in the language of physics, and therefore specifically, that every psychological proposition can be translated into an expression in which physical concepts alone occur.
>
> - Schlick, 1935c, [1979b, p.426]

Schlick therefore proceeds to investigate whether we can make sense of the idea that there is some psychological domain of facts which cannot be expressed as a function of physical language. He considers two possible ways the world could have been, one actual and one hypothetical. In the actual world, phenomena are consistently correlated with physical conditions. A specific shade of red is always associated with a certain wavelength of light, and if a person reports seeing a different shade of red in response to the same wavelength then a scientist will not be satisfied until they have found some irregularity in the condition of the subject, e.g. colour blindness. In the hypothetical case, there are no physical facts consistently correlated with certain phenomena, so we cannot reliably say "such-and-such shade of red relates to such-and-such wavelength of light". Of course, a certain amount of regularity in experience is required for us to be able to interact with the physical world at all, so in Schlick's examples the phenomena which are

freed from causality are relatively local. He envisages situations in which the entire world appears to be tinted red or blue or yellow, with the tinting shifting over time at random. In this case we would still be able to use machines to measure the wavelength of light etc., but there would be some phenomenal quality that wasn't related to that wavelength and we would be unable to predict under any conditions precisely what colour we would experience.

It might seem as though in describing these cases, Schlick is describing two worlds identical in form but differing in the content of their inhabitants' experiences, and thereby at least violating the incommunicability of content in his description, if not within the worlds themselves. I will note at this point, it is certainly *not* the case that incommunicability of content is violated *within* the worlds. An important feature of Schlick's hypothetical case is that the inhabitants of the world will not be able to speak to one another about their changing phenomenal states, because the act of speaking about it would create a physical fact – *propensity to respond in a certain way to questions about current experience* – that correlated with the phenomena. The physical language remains the only intersubjective language in the hypothetical case, but there are additional facts about colour that we can only reflect upon in a private language (Schlick, 1935c, p.433).[18] But if the case can be described at all, perhaps that is enough evidence to show that Schlick doesn't still think content is incommunicable while he's writing? Well, no, I don't think so.

We can ask ourselves, what would the subject be saying to themselves in the private language envisaged by Schlick that allows them to reflect upon their colour experiences? They could be saying that the world is still tinted the same way now as it was a few minutes ago, but it's slightly darker than how it was yesterday. They might say that this tint is the same tint as they had last week. All these things are purely formal relations between tints. As before, one doesn't *say* anything unless one relates two things together – that's what separates *kennen* and *erkennen* – so at

[18] We'll be discussing Schlick's use of private languages in much more detail in chapter 5, but for now suffice it to say that he thinks they're possible.

no point does the subject express the content of their experience to themselves. When they name the colour of their world in any moment, they merely relate it to other times when the world has shared this property. And when Schlick describes the hypothetical world, he too is only picking out the form of their experience. He is describing it as uniform at any given time and varied across times, and this uniformity and variety is related to the variety of colours that we are able to talk about in our own intersubjective language. That is to say, he is using our own intersubjective language to describe a world in which some aspects of our language could not be used intersubjectively (in essence, he described a world in which everyone is colour blind so that the people wouldn't have words for colours), but at every point – both in his description and in the hypothetical subject's private language – the language is formal. He is describing how experiences relate to one another, not how experiences are in themselves, whatever that would mean.

The other example Waismann mentioned is also picked out by Oberdan (1996, pp.281-284) as evidence of Schlick's changing position. It comes from "Meaning and Verification" (Schlick, 1936), and seems to me to be an even clearer continuation of the principle that Content is incommunicable. Schlick considers what might be meant by a solipsist putting forth the thesis "I can feel only my pain". We can compare a number of claims in a few different imagined scenarios:

(P) 'I feel pain only when body M is hurt.'

(Q) 'I can feel only my pain.'

(R) 'I can feel somebody else's pain as well as my own.'

(S) 'I can feel pain in other bodies as well as my own.'

Assuming 'M' refers to the speaker's body, in the real world P is true, but we can imagine scenarios in which we feel pain when other bodies are hurt and S is true instead. Suppose the speaker has a twin, N, and, in addition to feeling pain whenever the body of M is hurt, they also feel pain whenever their twin is hurt. The speaker's pain experience is the same when body M gets a paper cut on its thumb as it is when

body N gets a paper cut on its thumb, so when it happens they will have to visually check which thumb is bleeding. In this scenario, is Q true for the speaker?

Schlick argues that there are two ways of interpreting Q. On one interpretation, "*my* pain" is taken to refer to the sensation that accompanies damage to "*my* body". In this case, our imagined subject doesn't only feel his own pain, because he also senses damage to his twin's body. On this interpretation, P and Q are just different ways of expressing the same proposition. In the real world, both are true. In the hypothetical scenario, both are false, to be replaced by R and S. This means that Q is an empirical hypothesis – we understand what conditions verify or falsify it. But, says Schlick, this is not something a solipsist will accept. They are putting forth a philosophical view, not an empirical claim, and therefore the "can" in Q should be of a logical kind. It should be impossible to imagine a case in which a subject feels someone else's pain, and so, even when they can feel pain in the body of another, we must say that the feeling itself remains theirs. On this second interpretation, *my pain* is not necessarily linked to damage to *my body*. Thus, in the hypothetical scenario where P is false, the solipsist will reject R, continue to maintain Q, and accept S. So, on the first interpretation either P & Q or R & S will be true depending on the facts of the world; on the second interpretation Q is always true, R is always false, and only P and S have truth functions that depend on the facts.

Schlick argues that what this means is that, for the solipsist, Q is a tautology. If a solipsist uttered "I feel my pain" then the "my" would be redundant, because they have endorsed a rule of grammar on which the only kind of pain it is possible to feel is one's own, and so it says no more than "I feel pain". R, in contrast, is nonsensical, because there is no way to understand the claim that someone feels someone else's pain. "A tautology, being the negation of nonsense, is itself devoid of meaning in the sense that it does not assert anything. … the second interpretation of Q, adopted by the solipsist and forming the basis of his argument, is strictly meaningless," (Schlick, 1936, [1979b, p.477]).

Does anything about this example show that Schlick thinks the content of experience is communicable in 1936? He thinks we can make sense of someone

feeling pain when another person's body is hurt, but that isn't a case of shared content between two subjects. What's shared is an understanding of a relation between bodies and damage – it's the physical facts of the matter, not anything like the quality of experience. Schlick does think that it's nonsense to assert that there is something like Content that one person experiences and another cannot (let Q* be "I can experience only the content of my experiences" and run the same argument through again), but he already asserted as much in F&C. He made it very clear that it was merely a tautology to say that Content was incommunicable, and he argued that in the first lecture of F&C he hadn't actually asserted anything at all. There is nothing in the 1936 example that conflicts with Schlick's 1932 account, and there is even some reiteration of the fact that the meaning of a statement is unconnected to the phenomenal content that we associate with it:

> In order to understand a sentence containing, e.g., the words 'red flag', it is indispensable that I should be able to indicate a situation where I could point to an object which I should call a 'flag', and whose colour I could recognize as 'red' as distinguished from other colours. But in order to do this it is *not* necessary that I should actually call up the image of a red flag. It is of the utmost importance to see that these two things have nothing in common.
> - Schlick, M., 1936, "Meaning and Verification", [1979b, p.470], original emphasis.

So, it is quite clear that for Schlick the definition of "red" is not in any way fixed by the image of redness that we might have in our heads (or equally by what we see when we look at red things). To be a competent language user requires that we can distinguish red things from blue things, but not that we have some particular content associated with the word "red". His view on this matter seems to me not to change from 1932 to 1936.

So, what can be made of our earlier objection that Schlick seems to be attempting to say something ineffable? One of Schlick's earliest examples to illustrate the Form/Content distinction (F&C, p.291) was that of a gramophone disc,

which is a material that can express any sound (and therefore any sentence) depending on the structure of the grooves. Anything that can be expressed can be expressed by the gramophone disc, including every possible fact about the disc itself and the manner in which the grooves should be interpreted. But the disc cannot be expressed, because it is not the kind of thing that expression works with. The disc is not a fact, so the disc cannot be expressed, but the disc can take on a form which expresses a fact, and there are facts *about* the disc that can be expressed. Likewise, the content of an experience cannot be expressed, but this does not imply that we cannot express facts *about* Content. Schlick did write "there is no proposition about content, there cannot be any. In other words: it would be best not to use the word 'content' at all," (F&C, pp.306-307) but we can read this as a reiteration of the fact that no proposition can contain content, i.e. nothing the gramophone disc can say will involve the disc itself (you won't see vinyl shooting out of the speakers), and as a consequence of the fact that Schlick thought a language could not express its own grammar. Schlick took it to be a tautology that the content of an experience was inexpressible, which is to say that "the content of an experience is inexpressible" was a rule of grammar and not a meaningful proposition, but if we allow that languages *can* express their own grammar then propositions which involve the word "content" are possible (although not propositions which contain content) just as the gramophone disc can express statements about itself. So, it's fine for Schlick to write an essay expressing the distinction between Form and Content, so long as he doesn't try to express any content in it. And he doesn't - at no point does Schlick attempt to express anything like 'the greenness of green'.

Once you remember that rules of use are effectively rules of translation (see section 3.2.4) it becomes much easier to understand what Schlick is talking about when he uses the term "content". Consider a man looking at a red ball on a table. The fact which we are considering for this example is *there is a red ball on the table*. Before we get to considering the man, we note the fact that light travels from the ball to his eyes carrying, so to speak, the fact *there is a red ball on the table*. If we knew everything there was to know about the light on its way from the ball to the eyes then first and foremost we would know the facts about the structure of the light, but

we would also be able to apply certain rules to this structure and translate it into the fact *there is a red ball on the table* (or *it appears that there is...,* which is what the man will actually get from all this). If, on top of those facts, we also knew everything there was to know about how eyes respond to light then we would be able to apply rules to the facts about light and turn them into facts about the nerve signals the man's eyes will generate when they detect the light. Once again, facts about the structure of a particular material can, by application of certain rules, be turned into facts about the structure of a completely different material. The logical structure of *there is a red ball on the table* is present in the light and in the nerve signals from the eye so long as the structures of those materials are interpreted in accordance with the correct rules. The man's experience of the red ball on a table is one more link in the chain - the nerve signals from the eyes are interpreted by the brain according to certain rules in the same way as the light was interpreted by the eyes according to certain rules, only this time instead of translating light to nerve signals the nerve signals are translated into an experience. The content ***redness*** is the material that the brain translates the nerve signals into, and that ***redness*** is given a structure to match that of the ball (i.e. it is round and in a spatial relation with some ***brownness*** which represents the table). And now it should be obvious why it makes no sense to ask if another person has the same experience of redness as you. If one thing is the same as another it can either be the same in terms of form (as two identical tennis balls are) or it can be the same in terms of material (as a tennis ball is with its future self, even if in the future it is painted a new colour and therefore no longer has the same form as it did before). Obviously, another person's experience is not made out of the same material as yours - for that to be the case your mind would have to be inside theirs, which is a scenario that doesn't even make sense. But equally obviously, the other person's experience does have the same form as yours (assuming no hallucinations/impairments on either side) because agreeing on the logical structure of an experience is what it is for the two of you to agree on the colour of something. So, the question of whether or not two people have the same experience is shown to be trivial, the answer being obviously 'yes' or obviously 'no' depending on the interpretation of "same" being used.

3.3.5 A Problem for Structuralism

There is one more potential problem for the structuralist account of language which should be highlighted before we commit ourselves to the view, and that is the issue raised by Max Newman against Bertrand Russell (Newman, 1928). Once we've limited language to an entirely structural endeavour, our final system of science is going to say something like "∃x ∃y ∃z ∃R ∃S (xRy & ySz & …", which states that there are a number of objects that stand in a number of specified relations to one another but not what any of the objects or relations are. So far, so good – we've only outlined Carnap's account at this stage, but we've at least seen that a theory doesn't need to name the things which are being related in order to give a full account of the state of the world. The problem is, if we're liberal enough about what counts as a relation then that sentence will be true in any world which has the right cardinality. For any set of objects, we can define relations using arbitrary object-pairs which will satisfy the model. That means that a purely-structural description of the world doesn't say anything about the world beyond "this number of objects exists". That's a problem for a theory of science, which aims to say somewhat more than that.

The solution to the problem is going to have to be to put some limits on what counts as a relation. Schlick's account gives us a principled way of doing that, in that experience is necessary for teaching language to a subject so the relations that it is possible to name must be those which the subject is able to experience. This does, however, demand that *experience* remains a primitive concept in our system of science. We can't reduce the world entirely to logical vocabulary, as we will see that Carnap attempts to do in the next chapter. The Newman problem is therefore less of a problem for Schlick than it is for Carnap, but it will come up in the next chapter (see page 95) and play a big part in chapter 6, and, since it imposes restrictions on structuralist accounts, it seemed worth mentioning as we move to adopt Schlick's.

3.4 Conclusion

We have seen that Schlick's philosophy of language has shifted under the influence

of Wittgenstein to a picture theory, and that the meanings of words, although entirely structural, must be learned through experience. We have seen that Schlick's philosophy of Form & Content was not an attempt to express the ineffable, but merely an argument that a picture of a fact cannot be identical with the fact itself carried through to its logical conclusion. And we have seen that Schlick did not abandon major aspects of his philosophy in 1933 as some commentators have suggested – the incommunicability of content remains a feature of his philosophy for the remainder of his writings. In the next chapter we will step back and look at Carnap's position in greater detail. We will then look at the objection Neurath raised against it, which Schlick took to be a threat to empiricism itself.

4. Protocol Sentences

4.0 Introduction

While Schlick was developing his Form and Content philosophy, a debate was gathering pace within the Vienna Circle over how, precisely, statements about immediate experience were supposed to serve as a foundation for the system of science. The underlying problem is laid out clearly by Russell (1917, [1951]) in "The Relation of Sense-data to Physics", whom I will quote at length:

> What can we learn by observation and experiment? Nothing, so far as physics is concerned, except immediate data of sense: certain patches of colour, sounds, tastes, smells, etc., with certain spatio-temporal relations. The supposed contents of the physical world are prima facie very different from these: molecules have no colour, atoms make no noise, electrons have no taste, and corpuscles do not even smell. If such objects are to be verified, it must be solely through their relation to sense-data: they must have some kind of correlation with sense-data, and must be verifiable through their correlation alone. But how is this correlation itself ascertained? A correlation can only be ascertained empirically by the correlated objects being constantly found together. But in our case, only one term of the correlation is ever found: the other term seems essentially incapable of being found. Therefore, it would seem, the correlation with objects of sense, by which physics was to be verified, is itself utterly and forever unverifiable.
>
> – Russell, 1917, [1951, p.108]

That is to say, the objects of science, and of physics in particular, seem so far-removed from the objects of experience that it is unclear how experience could ever serve as justification for belief in the system of science.

Russell identifies two possible responses to the problem. One can either argue that we know *a priori* that our sense data have causes, or one can attempt to define the objects of physics as functions of sense data. In this chapter we will look at the Vienna Circle's attempt to take the second route by unifying the language of science with the language of experience. This would serve the dual aims of establishing that there was no separate sphere of psychological study that went beyond the study of physics, and, critically, that scientific knowledge was possible on the basis of experience. Of particular importance to this endeavour were protocol sentences – propositions about a subject's immediate observations. Initially these were taken (by everyone except Neurath) to be inherently justified, and they would be so related to the other statements of science that their justification was sufficient to justify the rest of the scientific system. Ultimately, however, Neurath's doubts about the justification of protocol sentences would fatally undermine this approach. The debate over the structure and status of protocol sentences would go on, in Schlick's eyes, to question the importance of experience for scientific inquiry, and this was what provoked his controversial work on the foundations of knowledge. We will start by looking at the role played by protocol sentences in Carnap's work, beginning with the *Aufbau* and continuing to their explicit introduction in "The Physical Language as the Universal Language of Science" (1932a [1934a]). We will then look at the problems raised by Neurath in "Protokolsatze" (1932, [1959]), and in the final section we will make explicit the position which Carnap and Neurath adopted to solve these problems.

4.1 The *Aufbau*

Protocol sentences were not explicitly mentioned in Carnap's *Der Logische Aufbau der Welt*, but it was through his work that they came to be of central importance to the philosophy of the Vienna Circle. In this section I will briefly talk about Carnap's goals and methodology in the *Aufbau*, and the role that protocol sentences played

there.[19] We will see the importance of the role that protocol sentences were to fill within Carnap's project, before moving on in the next sections to see how exactly they were to be fleshed out.

4.1.1 Carnap's Goals

We can distinguish between three readings of Carnap's goal regarding the reduction of objective facts to sense data (Pincock, 2005, p.519; Uebel, 2007, pp.43-51): the traditional reading (aka "phenomenalist foundationalism"), the so-called "neo-Kantian" reading, and the deflationary (aka "reserved") reading.

The Traditional Reading

On the traditional reading, Carnap wanted to pursue a radically anti-metaphysical project which would show that science was only committed to claims about sense experience. This would unify the languages of science in a single empirical language, taking totalities of experience as basic and abstracting qualities from them which can be recombined to give the meaning of terms in the language. We can see this reading in, for example, Ayer, who writes of the *Aufbau* as an attempt to justify the verification principle by showing that "statements about physical objects can be faithfully translated into statements about [subjective] sense-data, … a valiant attempt to reconstruct our whole apparatus of empirical concepts on a solipsistic foundation," (1959, p.13), and Quine, who says that the *Aufbau* "set itself the task of specifying a sense-datum language and showing how to translate the rest of significant discourse, statement by statement, into it," (1951, [2008, p.72]).

On this view, the project of the *Aufbau* failed when it attempted to bridge the gap between phenomenal sense data and physical objects - as Quine points out, Carnap provides no reduction of the connective "is at", which was supposed to position sense data within the intersubjective physical world. Since the whole goal, on this reading, was to analyse scientific concepts into phenomenal concepts, the

[19] For a more in-depth discussion of the issues in this section, see Uebel, 2007, pp.37-60.

failure of this step means the failure of the anti-metaphysical and foundationalist goals.

Uebel (2007, p.44; 1992, p.38) has argued that this reading cannot be reconciled with the *Aufbau*'s §106, where Carnap makes it clear that his constructive system is an example based on the current state of science, "in particular, upon the findings of the phenomenology of perception, and psychology". If the reduction of scientific knowledge to phenomenal experience depends on the results of scientific investigation into phenomenal experience then the reduction is circular (i.e. there would be an obvious problem with the report: "We find that people say 'I see red' when they see red, so when we say 'they see red' we mean that they say 'I see red'," - the initial observation could not be made without assuming the truth of the conclusion). Therefore, Carnap cannot have meant to reduce scientific knowledge to phenomenal experience. Uebel also points out that Carnap's emphasis on structure (which we will explore in more detail below) would have been unnecessary if he thought that knowledge could adequately be reduced to intuitive experience - recall from chapter 1 that Schlick's own attempt to reduce science to a logical structure was motivated precisely by the perceived inadequacy of experience as a foundation.

The Deflationary Reading

On the deflationary reading, put forward by Chris Pincock (2005), Carnap was not pursuing an empiricist project or a neo-Kantian project - he did not see himself as pursuing a philosophical project at all, but rather a scientific one. On this view Carnap's primary goal was to unify various systems as a single body of knowledge. Different systems which looked at things from different perspectives were possible. In the *Aufbau*, Carnap focuses on a system which organises things from an epistemic perspective. Objects are organised in order of epistemic primacy - some objects are recognised through mediation of others and so they are to be reduced to the objects by which they are recognised. For example, other minds are recognised by the behaviour of physical bodies, which in turn are recognised by an observer's phenomena, so other minds are reduced in two steps to an observer's phenomena. But, Carnap also mentions a system which takes physical things as basic

and constructs everything else out of them, which would be a system well-suited to picturing scientific knowledge. Pincock argues that the existence of this alternative system, which was worked-out in a 1925 paper and unpublished notes from around 1920 (Pincock, 2005, p.532), shows that Carnap was not committed to a foundationalist project and was merely interested in the interrelations between concepts from different areas of science. Physical concepts are still reduced to experience, but there is no reason to think that experience necessarily is the basic element for logical constructions - we could just as well reduce the phenomenal concepts to physical objects. The goal of the reduction in this case is to achieve unity of the languages of science - Carnap wanted to show that it was possible to translate, for example, the language used in psychology to the language of physics, and vice versa.

On either of the traditional or deflationary readings, Carnap is providing a system by which objective facts can be reduced to qualitative experience. Both readings regard this reduction as having failed, although for the deflationary reading this does not mean that Carnap's entire project must be abandoned - we merely need to use a different constructive system. The neo-Kantian reading, on the other hand, does not take Carnap to be attempting a reduction to experience at all.

The Neo-Kantian Reading

The so-called "neo-Kantian" reading, developed and named by Friedman (1987) and Richardson (1998), is not one on which Carnap was a neo-Kantian but one on which the *Aufbau* is read as a response to neo-Kantian problems – in particular the question of how objective knowledge is possible. On this view, Carnap's project in the *Aufbau* is first and foremost an attempt to explain the possibility of knowledge, and it centres on the structure of experience. Instead of reducing facts to experience they were supposed to be reduced to a particular structure. Experience would also be reduced to a structure, and it was this structure which was the most basic element of the system.

Carnap's notion of structure was supposed to be distinct from Schlick's implicit definitions, addressing the concerns which he had raised against them in

"Proper and Improper Concepts". Rather than defining a relation implicitly by the axioms in which it occurs, a structure description defines it by specifying which objects the relation holds between. Carnap gives the example of the relation *x is the father of y* in a three-member family, which he says would be implicitly defined by the following axiom system:

> 1. The field of R has three elements.
> 2. R and the R-chains are irreflexive.
> 3. R is intransitive."
> - Carnap, 1928, [1967, p.13]

In addition to the complaint we explored in section 1.2.5, that this structure can be laid over any number of things with the right interpretation and therefore shouldn't really be taken as definitive of a concrete concept like *x is the father of y*, Carnap raises the issue that there are even properties of the abstract structure itself that the axioms leave indeterminate. This axiom system has more than one possible realisation - it is satisfied in (a) the case where a man has two children, or alternatively (b) the case where a man has a son and that son has a child of his own. A structure description, on the other hand, would determine exactly one possible form that a realisation must take.[20] Suppose the family members are A, B, and C, and let (*x*, *y*) denote that *x* is the father of *y*. The two different cases mentioned above would then be represented as follows:

---

[20] One might object at this point that *x is the father of y* really *should* have two (or more) possible forms of realisation. The concept of *father* is obviously something we apply equally to men with one child or with many. But this would be to separate the relation from its field. It's a mistake to think of *x is the father of y* as being a one-one relation in some families and a many-one relation in others. Rather we should think of *x is the father of y* with regard to family P as being a different relation to *x is the father of y* with regard to family Q, so that there are two different relations, both of which we call "father", and which may have different structures in their respective families. Alternatively, we could think of a single relation *x is the father of y* across the field of all families which is neither one-one nor many-one, but which instead can only be adequately described with a complete structure description that told us about every father. The various possible concepts of *father* thus constructed may not line up with the intuitive meaning of the term "father", but Carnap was explicitly aiming for extensional - not intensional - adequacy (1928, [1967], §51).

(a) (A, B) & (A, C)

(b) (A, B) & (B, C)

Carnap also mentions the possibility of expressing structure descriptions with arrows between points, with each point representing an object and the arrow indicating the existence of a relation between that object and the thing it points to (which may be back to itself in the case of a reflexive relation) (1928, [1967, §11]).

Carnap held that structure descriptions picked out proper concepts, as opposed to the improper ones of Schlick's implicit definitions, in part because they described all of the properties of the relation and left nothing indeterminate. In the case of *x is the father of y*, for example, rather than one axiom system which has two possible forms of realisations we have two structure descriptions which each have only one form of realisation. With relations thus completely described by their structure, individual objects can now be picked out by definite descriptions based on which relations they did and did not satisfy (1928, [1967, §13-15]). Carnap acknowledged that this might not be possible in all possible worlds, but if there are a sufficient number of objects related to one another with a sufficiently variegated structure then it can be done. He takes it as a hypothesis of science that there is in fact a sufficiently complex world to pick out objects from their relations. And if two objects cannot be distinguished by structure then, Carnap says, science must take them as perfectly similar (i.e. qualitatively, if not quantitatively, identical). This picking out of objects could not be accomplished on the model of implicit definitions, because when the structure is not completely determined we can't use that structure to pick objects out unambiguously.

On the neo-Kantian reading, Carnap's project consisted in reducing scientific knowledge to structure. It was this structure, rather than phenomenal experience, which was to be objective and intersubjective, and it was the structure rather than the sense data that one had to grasp in order to have grasped scientific knowledge. This reading makes the *Aufbau* seem very similar to what Schlick was later going to suggest in "Form and Content", which we have already looked at. Objective

knowledge involves stripping the content away from our experiences and focussing on the logical form.

The problem which arises for Carnap's project on the neo-Kantian reading is what Uebel calls "the foundedness problem". The issue is that the structural reduction appears to be incomplete. "A purely structural statement must contain only logical symbols; in it must occur no undefined basic concepts from any empirical domain," (Carnap, 1928, [1967, §153]), but the basic relation with which Carnap constructs the rest of the concepts of science (recollected similarity; Carnap, 1928, [1967, §108]) never gets structurally defined. Carnap's attempt to eliminate the basic relation from the system involved an appeal to experience - the idea was to reduce statements involving the relation to a set of ordered pairs and let the relation be defined as that relation by which these pairs demonstrated some kind of empirical connection (it being highly unlikely that in a complete system there could be more than one relation which held between exactly the same pairs of objects). This meant that the basic relation in any constructional system had to be empirically grounded (or "founded") in some way. But this just brings intuitive experience back into the picture, which (on this reading of Carnap's goals) is precisely what we were trying to avoid. The problem of the re-introduction of a concept of *experience* is related to the problem we highlighted for Schlick's structuralism in the last chapter, and it will return when we come to discuss David Chalmers' recent re-evaluation of constructive projects in section 6.4, under the name of "the Newman problem".

Sub-Conclusion

Uebel concludes (and I concur) that both the neo-Kantian and the deflationary readings capture elements of Carnap's motivations:

> That in the *Aufbau* Carnap accepted as relevant the foundedness problem shows that he was driven by more than a deflationist concern to explicate objectivity from within; that later he accepted the relevance of Quine's criticism also shows that the phenomenalist reduction was not just pursued as a convenient example of rational reconstructionist philosophy. So both concerns

> with the conditions of the possibility and with reduction were driving the Aufbau.
> - Uebel, 2007, p.57

The project of the *Aufbau* was to explain how objective knowledge was possible, and to show how scientific concepts could, in principle, be constructed out of given experience. The goal was *not* to show how our concepts are *in fact* constructed - that was a task for psychology, and, as we saw Schlick point out before (AE, p.47), there is likely to be more than one way to construct any given concept. In Carnap's words:

> The old concepts did not ordinarily originate by way of deliberate formulation, but in more or less unreflected and spontaneous development. The new definitions should be superior to the old in clarity and exactness, and, above all, should fit into a systematic structure of concepts.
> - Carnap, 1928, [1967, p.v]

Besides, the question of which particular route was taken is not as important as the question of whether or not any routes exist. Any reduction will show that it is possible to unify the languages of science.

This means that Carnap's project can achieve partial success by reducing objective knowledge to the structure of experience, but we will need to attempt a further reduction to phenomena if we are to achieve complete unity of the language of science.

4.1.2 Logical Constructions

The process of constructing a concept consisted of showing how statements about the concept could be translated into logically equivalent statements which did not mention it. Carnap gives the example that fractions can be constructed out of natural numbers, which can be seen by translating a statement about, say, $\frac{2}{3}$, into a statement about 2 and 3 (§2). The system which Carnap develops in the *Aufbau*

attempted to construct all of the concepts of science out of sense data and the relation of remembered similarity between them.

Carnap took experience states as his starting point because, although a construction which took physical objects as its starting point would be possible (p.vii; §59), part of his goal was epistemic. Experience seems to be a natural starting point for an epistemic project (§64) - it seems like we get our evidence from experience in the first instance and derive the rest of our knowledge from there. The goal is to construct a system which links our experience with objective, intersubjective knowledge claims:

> Even though the subjective origin of all knowledge lies in the contents of experiences and their connections, it is still possible, as the constructional system will show, to advance to an intersubjective, objective world, which can be conceptually comprehended and which is identical for all observers.
> - Carnap, 1928, [1967, §2]

This meant that at some point Carnap was going to have to separate the subjective elements of experience - what we have called "content" - from the objective logical form that the experience conveys. This is the job which would eventually be performed by protocol sentences - the protocol sentences themselves were to be constructed in a private language and refer to experience, but they would be translated into the physical language in a way which preserved their logical form.

Carnap did not highlight the role played by protocol sentences in the *Aufbau* (indeed, the term "protocol sentence" was coined by Neurath some three years after the *Aufbau* was published (Uebel, 2007, p.180)), but it is with regard to the method of logical reconstructions that they gain their importance. With all the sentences in the system of science translated into sentences about experience, the form that these sentences take and the epistemic status they hold is of the utmost importance. In the next section we will look at how Carnap developed and clarified his position on protocol sentences over the years from 1928 to 1932.

4.2 The Physical Language as the Universal Language of Science

Carnap's explicit development of protocol sentences roughly coincides with the syntactic turn of his philosophy. This was the time in which he came to believe that all of the significant features of a sentence, including its truth value, could be found in its syntax - a view we briefly contrasted with Schlick's position in the previous chapter.

By the time of "Physical Language..." (1932a, [1934a]) Carnap had been persuaded by Tarski of the importance of the distinction between object language and metalanguage (Uebel, 2007, p.142). This comes out explicitly in Carnap's distinction between the material and the formal modes of speech, which Carnap runs alongside each other in separate columns. The material mode of speech talks about objects and states of affairs, while the formal mode of speech talks only about words and sentences. For example, we might say in the material mode of speech that arithmetic is the study of the properties of numbers, but in the formal mode of speech we would say that sentences of arithmetic are made up of signs of such-and-such a kind, put together in such-and-such a way, and can be transformed according to such-and-such rules. Carnap believes that "a philosophical investigation must be an analysis of language" (1932a, [1934a, pp.37-38]) and so the formal mode of speech is, strictly speaking, the correct one. The material mode of speech employs statements which only make sense within the object language, but the formal mode of speech employs statements in the metalanguage.

Carnap draws attention to three possible views that one could have about the nature of protocol sentences, and, in 1932, does not attempt to choose between them.[21] On the first view, protocol sentences are of the form "here now blue, there red", and the objects of protocol statements are, in the material mode of speaking,

[21]In “Uber Protokollsatze”, written after Neurath’s criticisms and published at the same time (still in 1932), he argues that the three possible views are actually compatible with one another and it’s just a matter of practicality that we might choose to use one over another.

the "simplest sensations and feelings" (p.46), not whatever physical things may or may not have caused them. On this conception, the word "blue" in the protocol sentence "here, now, blue," does not refer to some property of some object in some external world - it refers to the speaker's mental state at the moment of utterance. Carnap attributes this position to Ernst Mach.

The second view is that what is actually recorded in a protocol sentence should be some complex of current experience. Depending on which proponent of the view you were speaking to, this could mean either some partial region of one's perceptual (e.g. visual or aural or ...) field, the entirety of that perceptual field, or the entirety of experience across all perceptual fields at the same time. Sentences like "here, now, blue," can be derived from protocol sentences by isolating a certain part of the speaker's current experience, but they are not strictly speaking protocol sentences themselves. This is the view which Carnap says was held by the majority of people at the time. The distinction between this view and the first one is that experiences at individual points – the experiences which would be described by “here now red” for instance – have to be isolated from the given protocol. This view was adopted by many positivists in response to results in Gestalt psychology which appeared to show that our experience is something other than the sum of many individual points (Carnap, 1932a, [1934a, p.47]).

The third view is that protocol statements have essentially the same form as statements like "the red cube is on the table". On this view, the objects of protocol sentences are objects of the external world. A protocol sentence is not a sentence about one's direct sense experience, but about what it is that one is currently experiencing. Carnap says that this view was not widely held at the time, but we shall see in the next section that it is (at least, close to) the view that was soon to be advocated by Neurath. That Carnap was willing to entertain this view as an option is already a step towards the view he would eventually adopt - in the earliest draft of the paper, from around mid-1930, the only options considered were the phenomenal ones (Uebel, 2007, p.188).

However we think of protocol sentences, Carnap says that they must be translatable into sentences of the physical language - that is, the language which expresses what properties of the world are in what place at what time. The protocol sentences themselves were to be formed in a separate language, called the "protocol language". The physical language is the language of science, and because science is not meant to have "merely subjective interpretation but sense and validity for all subjects who participate in it" (1932a [1934a, p.66]), the language must be intersubjective. That means that words like "red" in the physical language don't correspond to a specific private sensation (as they do on the most common interpretation of the protocol language), but to properties in the physical world like "the colour at such-and-such position on the colour chart". If there was a machine which spoke out loud any colour at which it was pointed then a blind man could work out where certain colours were and could use the words "red" or "blue" in the physical language, even though they would not appear in his protocol language because they could only ever be obtained indirectly.

In addition to being intersubjective, the physical language is supposed to be universal. That is to say, "every statement (whether true or false) can be translated into it" and "every possible state of affairs (every conceivable state, whether actually occurring or not) can be expressed by it" (Carnap, 1932a, [1934a, p.67]). If a statement cannot be translated into the physical language, then it is not verifiable and is therefore a pseudo-statement. That includes protocol statements. Carnap says that "every statement in the protocol language can be translated into a physical statement and indeed into one which describes the state of S's body" (1932a, [1934a, pp.87-88]) - this link between the epistemically privileged sentences in the protocol language and the sentences in the physical language was supposed to be the epistemic foundation of the system of science. However, this step introduced a problem. Protocol sentences were supposed not to require verification (or, at least, they were supposed to be undoubtedly true and immediately verified for whoever was formulating the protocol sentence) but any statement about the state of your body will be open to disconfirmation - an ardent sceptic could even maintain without contradiction that you might be a disembodied consciousness undergoing

utterly unreal hallucinations. It is this weakness that Neurath attacked in *Protokollsatze* (1932, [1959]).

4.3 Neurath's Formalisation of Protocol Sentences

The first step in Neurath's argument is to argue that a protocol sentence cannot merely be along the lines of "here now blue". Rather, he says, all protocol sentences must include a proper name. Protocol sentences always express the experiences *of somebody*,[22] and sentences in the language of science couldn't contain indexicals or demonstratives like "I" or "here" without losing their intersubjectivity. That means that a complete protocol sentence will read along the lines of "Otto's protocol at 3.15: [at 3.14 Otto said to himself: (at 3.13 there was a table in the room perceived by Otto)]" (Neurath, 1932, [1959, p.202]). The innermost set of brackets in the protocol statement contains a perception which is attributed to a specific person, and importantly the perception is intersubjective – it is perception of a physical fact rather than a phenomenal description of the perception (in other iterations of the proposal, Neurath suggested an additional set of brackets so that the innermost was a straightforward factual statement like "there was a table in the room" and the next layer out was something like "at 3.13 Otto perceived:" (Uebel, 2007, p.379)) . We could construct protocol sentences without the set of brackets containing "Otto said to himself", (e.g. "Otto's protocol at 3.15: [at 3.14 there was a table in the room perceived by Otto]") but it is useful to Neurath's argument to note that we can use sentences of the same form as protocol statements to describe people's experiences when they are hallucinating or lying (e.g. "Otto's protocol at 3.15: [at 3.14 Otto wrote: (at 3.13 there was a square circle in the room perceived by Otto)]").

[22] This is in stark contrast to Carnap's position in the *Aufbau*, where he had an entire section titled "The Given Does Not Have a Subject" (§65) about why experiences did *not* have to be linked to an individual. We will return to Carnap's position on this matter in section 4.4.1, as there is some scope for interpretation on this point.

Having introduced this formal structure Neurath points out that protocol sentences are statements of fact. And, like other statements of fact, they are subject to disconfirmation. Neurath asks us to imagine a scholar who can write with both his right and left hands at the same time. Suppose that the scholar writes a protocol sentence with each hand at the same time, and suppose further that the two sentences directly contradict one another. Clearly the two sentences cannot both be true. Therefore, protocol sentences cannot be sentences which require no confirmation.[23]

Neurath reiterates a point made by Carnap that science consists in forming a non-contradictory system of protocol and non-protocol sentences. However, since protocol sentences are just as subject to revision as non-protocol sentences, if we are faced with a protocol statement which conflicts with our system as it stands then we are not forced to reject the system - we can reject the protocol sentence. "A defining condition of a sentence is that it be subject to verification, that is to say, that it may be discarded," (Neurath, 1932, [1959, p.204]). Since no sentence is held to be true on its own, on Neurath's view it appears that the truth of a statement consists in nothing more than the mutual agreement of that statement with all of the other statements in the system. This is what is known as a "coherence theory of truth", and it is as opposed to a "correspondence theory of truth" under which the truth of a statement consists in it accurately reporting facts. The coherence theory of truth

---

[23] At first sight this argument seems so obvious that it must have missed the point somewhere. Even without the preceding discussion about the form of protocol sentences, it seems clear to us that for any sentence about immediate experience there is a similar sentence about a different experience mutually exclusive with the first. Clearly, we cannot work out just from looking at one such sentence whether or not it is true. Neurath's argument appears to have gone unnoticed for so long because the Vienna Circle did not think of propositions in the same way as we do today. Instead of there being infinitely many propositions, any of which a given utterance may or may not assert, they would have said that for a proposition to exist was for it to be asserted. Hence the contradiction only arises when someone *actually asserts* two conflicting propositions at the same time, as with Neurath's ambidextrous scholar. In fact, the outermost set of brackets, “Otto’s protocol at 3.15…” is considered satisfied only if what follows is in fact entered into the scientific record (Uebel, 2007, pp.383-384). It is an important part of Neurath’s philosophy that science is a social endeavour and that coherence is to be strived for within the sets of statements that people are in fact making, not merely within any arbitrary group of possible statements.

is, as we shall see, at the root of Schlick's complaints against Neurath, although it may not be attributed to him fairly (which we will return to from section 4.4.2 onwards).

Neurath argues that we should abandon any notion of a difference between our own protocol sentences and those of anyone else. It might be thought that we could be more certain about our own experiences than those of anyone else, but Neurath points out that memory is fallible, and for all intents and purposes a protocol sentence we wrote yesterday might as well have been written by someone else. We can imagine, for example, a person losing his memory and rediscovering his life by reading through a list of previously written protocol sentences, but we can just as well imagine him learning the history of someone else's life by doing so. Hence, we need an intersubjective protocol language for thinking about our own experiences just as much as we would need one for talking about someone else's. (This argument pre-empts the private language argument later attributed to Wittgenstein by some twenty years.)

Neurath's arguments showed that protocol sentences couldn't be formed in a private protocol language before being translated into the physical language, as Carnap had envisaged them. They would have to be part of the physical language to begin with. But, as part of the physical language, they lost both their connection to phenomenal sense data and their privileged epistemic status. In the next section we will look at the positive proposal which Neurath and Carnap built to address these problems, and in particular we will look more closely at the coherence theory of truth which they adopted.

4.4 Radical Physicalism

In this section we will develop a fuller account of the position adopted by Neurath and Carnap around 1932.

4.4.1 The Rejection of Methodological Solipsism

As mentioned in the previous section, Neurath rejected the idea that there might be anything to gain by talking about one's own protocols as opposed to those of another. "The 'methodological' solipsism and 'methodological' positivism do not become any more useable by the addition of the word 'methodological'," (Neurath, 1932, [1959, p.206]).

Neurath has a number of arguments on this point. His first is again reminiscent of Wittgenstein's later argument against private languages. He says that if a man wants to incorporate his present experience into his own system of beliefs then he will need to compare his present experience to his prior experiences. But his prior experiences, not being present, must make use of an intersubjective language. We cannot compare our experiences directly to one another - we can only compare to our memories, which are of a different material. And if we were to lose our memory and relive our experience by reading a diary or some other record, then there would be absolutely no difference between our own record or that of another.

Neurath also suggested that Carnap had failed in the *Aufbau* to provide an adequate account of how the autopsychological language was to make intersubjectively verifiable claims. There are two passages of the *Aufbau* which address intersubjectivity - §§65-66 gives a broad overview of the approach and §§145-149 give a specific example of how to proceed. The two passages can be read as giving different accounts (Uebel, 2007, pp.132-133), but they are not mutually exclusive. In §66 Carnap describes something like the method familiar to us from our discussion of "Form and Content" - "even though the *material* of the individual streams of experience is completely different … certain *structural properties* are analogous for all streams of experience" (Carnap, 1928, [1967, §66]). What's common to different observers of the same phenomenon in this model is the logical structure of their experience, and it's that which gets communicated in expressions and which is incorporated into the system of science. In §146, on the other hand, Carnap describes a subject constructing the experiences of another mind by analogy. He says that the system of a person M (S_M) will be a subsystem of the subject's

system S, treating objects as intersubjectively real if the other person in fact agrees to all of the same statements about the object as the subject. Uebel suggests that these approaches are strikingly different - one assumes the existence of other subjects and one does not - although he admits that they could be combined if we were to say that the consistency between S and S_M was the result of structural similarity between those worlds. Uebel thinks that this interpretation would cause further problems, but it is irrelevant for our exposition here.

Neurath's argument against Carnap's purported intersubjectivity is that agreement with a solipsistically constructed other does not, by itself, give us an objective standard of truth. If M is constructed within S then congruence between S and S_M is not intersubjective agreement - it's a kind of internal consistency. Structural agreement, on the other hand, whilst intersubjective, requires an intersubjective language and not an autopsychological one. Statements reduced as far as immediate sense data could not therefore be intersubjectively verifiable. Either Carnap's objective-knowledge project fails or his phenomenological reconstruction does.

Neurath's proposed solution, and the one adopted by Carnap, was the abandonment of the phenomenological project. Protocol sentences were to be constructed in intersubjective language and no longer needed to be indexed to the subject forming them. Removing sense data from the construction had the consequence that it was no longer clear how experience could determine the truth or falsity of protocol sentences. The consequence was the adoption of the coherence theory of truth (or more properly, of acceptance, as we will soon see).

4.4.2 Coherence

We have seen before how Schlick had always maintained that the meaning of a sentence consisted in its unambiguous association with a particular state of affairs, and the truth of a sentence consisted in that state of affairs actually obtaining. We also saw how this view was further developed by Wittgenstein's idea that sentences could picture the world by matching its logical form in a new material. The idea that the truth or falsity of a sentence consists in a corresponding state of affairs obtaining

or failing to obtain is known as the correspondence theory of truth. It is a view which Neurath rejected. In its place most people at the time thought he was advocating a coherence theory of truth, which is the position that would be attacked by Schlick, but we will see that he is better understood as bypassing the notion of truth and advocating a coherence theory of justification.

Neurath criticised the correspondence theory of truth in his earlier work, "Ways of the Scientific World-Conception" (1930, [1983]), in which he argues that there are times when two mutually exclusive models work equally well and that in those cases there is no one true model which properly corresponds with the state of the world (1930, [1983, p.45]). Separately, he makes the point that all we ever have access to are our own thoughts and there is no way to independently check whether or not our thoughts are true. We can only evaluate them with regard to our other thoughts, and the only standard we can aim for is coherence (1930, [1983, p.46]).

Science is nothing more than a system of mutually consistent statements. When we are faced with a new statement about experience that cannot consistently be placed within the system of science, we have a choice. The system must remain consistent, so we can either reject the statement or we can amend other statements within the system to accommodate it. We have seen that Neurath argued that it was impossible for statements about immediate experience to occupy a special place in the system of science. Thus, he argued, there was no way to decide in advance whether the sentence about experience or the rest of the system would be rejected. There was no assumption that protocol sentences would be true - the only standard was coherence within the system.

There are at least two ways of reading Neurath's position on this matter. The first that we shall look at was how Schlick understood him, which informed his critique of Neurath and which I think it's fair to call the naïve interpretation. This reading is important for understanding why Schlick reacted as he did and what he was trying to achieve with his later arguments. The second interpretation – which I will call the charitable reading - is more reasonable, and it is important for assessing the philosophical merits of Schlick's later position.

4.4.3 The Naïve Reading

On the naïve view, Neurath seems to be saying that there is nothing which determines the truth or falsity of a set of statements beyond their mutual consistency. Science is an internally consistent system of statements written in the language of science which is intersubjective - immediate subjective sense data don't enter into it. Protocol statements are a subset of those which record the immediate experience of individuals, but which otherwise have no special status within the system. It is an empirical fact that individuals believe certain protocol sentences, particularly ones which record their own experiences, and it is up to us whether or not we incorporate those sentences into the system of science. If they are consistent with the system and there is no special reason to doubt the subject, then they are likely to be incorporated. If they are not consistent, for example if someone reports experiencing an object falling in a vacuum at $2m/s^2$, then we have the choice of amending or rejecting sentences already within the system, for example the strength of gravity or the mass of the earth, or rejecting the protocol sentence (perhaps incorporating an alternative sentence like "this subject is an unreliable witness"). The decision we make at this point is not governed by rules of science - it is likely to be determined by a combination of social and historical factors and the elegance/simplicity of the resulting system. When a person says they experienced something that doesn't fit with the established system of science, we are likely to conclude that they were dreaming/hallucinating, but an equally viable solution would be to say that the rest of us have been hallucinating the whole time and this one individual is the only person who sees things as they really are. And if we made that decision then what that individual saw would be true (but this only sounds ridiculous because we know we would not make that decision). Truth consists entirely in the coherence of the system.

Schlick saw this approach as "dogmatic and rationalistic" (1935a, [1979b, p.400]), as well as believing that Neurath and Carnap had abandoned empiricism by severing the link between experience and truth. He criticised it as allowing “any fabricated tale to be no less true than a historical report,” (1934, [1979b, p.376]) so long as one takes care to construct the fiction without internal contradiction. It

seems ridiculous to imagine that a reasonably intelligent person (such as, I think we may assume, Neurath was) would have advocated such a view, so let's turn to the charitable interpretation.

4.4.4 The Charitable Reading

The problem that we are considering is how to close the gap between sense data and the system of science. The approach which Neurath rejected was defining the terms of the system of science as functions of sense data – staying on the sense-data-side of the gap and bringing the science to us. It appeared as though he was trying to cut sense data out of the picture entirely, simply constructing whatever systems of science would stand on their own and leaving experience as irrelevant to the process, but this allows us to believe whatever we like about the world and surely can't have been the position he was trying to advocate. On the more charitable reading, Neurath was actually trying to deny the existence of a gap in the first place.

As before, science is a system of mutually consistent statements written in the intersubjective language, of which protocol statements are a subset. Protocol statements do not occupy an epistemically privileged position, that is, they are subject to confirmation and disconfirmation in the same way as any other statement in the system, but they are nevertheless important for epistemology and distinguished from other statements in virtue of their unique form. Neurath writes in a letter to Carnap "If in response to the question 'How do I know this?' we end with a statement, then everything is in order. But no recourse must be taken to 'experience'; to ask of an observation statement 'How do I know this?' is meaningless." (RC 029-17-03 ASP, p.5, translated and referenced in Uebel, 2007, p.225). That is to say, protocol statements (or "observation statements" in the quoted passage) serve as the justification for other statements in the system of science, but the question of where they come from or how they are justified themselves is simply misconceived. They do, however, serve to ensure that the system of science remains grounded in experience, contrary to Schlick's misplaced concerns (Uebel, 2007, pp.386-387). Neurath's model crucially includes both a perception claim (in the innermost bracket, or second-from-innermost on the four-

layer model) and the requirement that the statement be asserted in the process of scientific inquiry (the outermost bracket), meaning that the system of science is to be based in (as far as we know, reliable) testimony about experienced facts. A protocol sentence that did not talk about experience would be ill-formed, and we must assume that in the actual process of science there is relatively little assertion of fantasies or falsified results.

On this reading, in contrast to Schlick and Wittgenstein's starting point that truth consists in the coordination of fact and proposition, Neurath was starting from a coherence theory of justification. Neurath's argument was that we never have an independent vantage point from which to compare protocol statements to the world they were supposed to picture, so when we come to a judgement about a statement it can only ever be by comparing that statement to other statements which we already assent to. As for whether or not these statements are true, Neurath appears to have thought that the question makes no sense. Later he even suggested that the words "true" and "false" could be removed from the scientific vocabulary altogether (Uebel, 2007, p.233). He kept using the terms only for the sake of "linguistic continuity", but meant by them whether or not a statement could be accepted, not whether or not its logical form was isomorphic with that of some part of the world (a position which he viewed to be meaningless metaphysics).

We have seen that Neurath's account of the acceptance of statements is not as ambiguous as Schlick took it to be – we don't get to accept just any consistent group of claims, however fanciful, on the grounds of consistency. Neurath's *conditions* for acceptance are clear enough, and do not admit Schlick's objection. What it *means* to "accept" a statement on Neurath's view, however, remains open. Generally, we would say that there is no difference between "I believe P" and "I believe that 'P' is true". If accepting a statement is not the same as thinking that the world is a particular way, then what is it? It seems like Neurath's position is that truth is simply irrelevant. Where Schlick and Wittgenstein were hoping to come up with a system under which scepticism is refuted, Neurath thinks that there is never absolute certainty for any proposition, so a direct link between justification and

truth is not part of his system as it was for Schlick. Neurath points out that justification in science stems from the agreement of statements with one another, not with subjective phenomenal sense data, and so there is never a need to connect statements to the phenomenal world with protocol statements or anything else. Neurath solves Russell's problem neither by constructing physical objects as functions of sense data nor by positing *a priori* knowledge of an external world, but by challenging Russell's assumption that sense data was involved in the process of verification.

4.5 Conclusion

The problem of the ultimate justification of statements in the system of science was a matter which provoked a great deal of heated debate within the Vienna Circle. We have seen how Carnap's attempt to ground the intersubjective sentences of science in a private protocol language failed at the point where he had to specify the spatio-temporal location of subjective data in intersubjective terms. The problems caused by private languages prompted Neurath to reject them altogether, and we have seen how he was led to a view on which verification was not a process of comparing propositions to experience, but one of comparing propositions to each other. This was a view on which truth and justification came apart, bringing back the scepticism which Schlick thought he had seen off in AE and, in Schlick's eyes, cutting the role of experience out of empirical science. Schlick saw this as unacceptable, and he responded in 1934 with "Über das Fundament der Erkenntniss" ("On the Foundations of Knowledge", henceforth "UFE") and the introduction of *Konstatierungen*.

5. Confronting the Real World

5. 0 Introduction

We have seen that throughout his career Schlick was focussed on the foundations of science and attempting to explain how we can establish them with absolute precision and certainty. The problem which Neurath identified, and which we investigated in the previous chapter, can be thought of as forcing a dilemma upon such attempts which seems like it will always render them unsatisfactory. On the one side, we may retreat to a private protocol language of immediate experience. This would leave our claims immune to disconfirmation but would also leave them unrelated to the intersubjective system of science, meaning they're not worth much. Nothing we could say could be incorporated into a system of the kind we want, and we could not even use our own knowledge from yesterday. On the other side, we can keep the relation to the intersubjective system of science but accept that our protocol sentences are potentially revisable. There are no certainties at the base of the system – epistemically it floats free, and in the face of inconsistencies the question of which propositions are to be revised is not settled in advance by philosophy. We have already explored the latter of these options, as it is the one chosen by Neurath himself and, eventually, Carnap. In this chapter, we're going to see how Schlick attempted to solve the problem through *Konstatierungen* (usually translated as "affirmations").[24]

In the first section I will introduce what Schlick says about affirmations and consider the various interpretations which have been given to them by recent commentators. Thomas Oberdan and Thomas Uebel each give interpretations on which one of Schlick's goals is achieved but the other is not – Oberdan provides an

[24] I will stick to the prevailing translation here, but I do think that Ayer's translation of the word as "confirmations" makes more sense in modern English.

account of affirmations standing in a clear relationship to the system of science but without the certainty that Schlick was aiming for, whilst Uebel suggests that we should think of affirmations as infallible statements which bear no deductive relationship to the system of science but might be able to provide psychological motivation for the acceptance of one or another set of protocol sentences. I will suggest that no interpretation could fully achieve both of Schlick's aims, but in the second section I will attempt to provide an interpretation which would explain why Schlick thought he could do so. There are important aspects of both Oberdan's and Uebel's interpretations which need to be preserved, such as the role of demonstratives and grammar in affirmations, but I will argue that the interpretation which makes the best sense of what Schlick was saying is heavily Wittgensteinian, in particular making use of the latter's notion of criteria. I will also note, however, that this doesn't seem to get Schlick everything he was looking for – it is probably the Wittgensteinian view that Schlick was attempting to express, but Wittgenstein's criteria don't appear to be capable of bridging the gap that Neurath had identified. I will suggest that only an essentially-solipsist account in which the system of science is interpreted in a private language could come close to achieving Schlick's aims, although this would of course be fraught with its own difficulties.

5.1 Recent Interpretations of Affirmations

As noted, Schlick intended to ground the system of science on individual moments of experience. He accepted that Neurath had shown that protocol sentences, interpreted as intersubjective statements made in the language of science, could not be held true come what may. However, he rejected Neurath's claim that therefore all statements were hypotheses. Beyond these basic objectives, the details of Schlick's proposal are obscure and often appear self-contradictory. In this section I will first consider Oberdan's interpretation, which attempts to rehabilitate Schlick's view into a coherent semantic system by ignoring the interest in certainty which motivated it. I will then look at how Uebel expounds the more traditional view, on

which affirmations are more like protocol statements in a private language which do not stand in a deductive relationship to the system of science. Lastly, I will consider whether there is room for a view between the two, and I will argue that there are good reasons to think not. Nevertheless, in the next section we will be exploring how Schlick might have thought that Wittgensteinian ideas could bridge the gap.

5.1.1 Oberdan's Interpretation – Maintaining the Link to Science

In "Postscript to Protocols: Reflections on Empiricism" (1996), Thomas Oberdan provides an interpretation of Schlick's affirmations on which "if Schlick's silliness about certainty is charitably ignored, the resulting epistemology represents a viable alternative to the physicalists' conception," (1996, p.271). Oberdan argues that, counter to the traditional view, the account of language that Schlick expounds provides "a tenable account of the determination of the truth of empirical statements, which does not rely on the doctrine of certainty" (p.271). I think that it would be so 'charitable' as to wildly misrepresent the text if we were to strip the certainty away from Schlick's philosophy. Nevertheless, Oberdan's account is worth looking at, both for the philosophical insight we might gain by stripping Schlick's views down for parts, and for illuminating the problems that arise when trying to give an account of affirmations.

Oberdan takes it that affirmations are not mental acts or events, but are a particular class of linguistic expression (1996, pp.272-273). When Schlick introduced affirmations in UFE he was not careful about drawing this distinction. Affirmations were said to "occur" (UFE, p.382) rather than being "formed" or "made", and Schlick speaks of our feeling of fulfilment when they occur, writing "These moments of fulfilment and combustion are of the essence. From them comes the light of all knowledge," (UFE, p.387). But at other points in the same essay Schlick says that they have the form "Here now so-and-so" (UFE, p.385), and he conflates them throughout with "observation statements" which explicitly are made rather than occurring. In Schlick's later work, "On 'Affirmations'", however, he discusses the grammar involved in their use – he claims that, unlike other

statements, affirmations would be rendered meaningless by the addition of bet-hedging words like “probably” or “maybe” – and says that "where *this* usage prevails, I call the statement an 'affirmation'", (1935b, [1979b, pp.409-410], original emphasis). On this reading, Schlick's talk of feelings of fulfilment relate more to how affirmations confirm hypotheses than to the nature of the affirmations themselves. It is perhaps worth noting here that UFE was written in a single draft one afternoon and may be a less careful expression of Schlick’s views than his later writings – several misquotations-from-memory of Carnap were cut only after the first proofs came back from the printers (Schlick, letter to Carnap, 13/05/34, referenced in Uebel, 2007, pp.301-302; Schlick, 1935a, [1979b, p.400]).

Once affirmations have been pinned down as linguistic expressions, we can start to see what role they might play in a theory. Oberdan writes that they are best understood as "protocols of an independent phenomenal language, distinct from the language of the scientific system" (1996, p.272). This approach to protocols was discussed by Carnap (1932b, [1987]) (see section 4.2), who concluded that in this case the phenomenal language must be translatable into the intersubjective language of science, otherwise affirmations could never fulfil their role of verifying scientific hypotheses. The system of science makes predictions in the language of science so the only thing which can confirm or disconfirm those predictions is the truth of a statement in the language of science or in a language which can be accurately translated into the language of science: if affirmations are to confirm predictions then there must be some logical relationship between them and the prediction. Schlick confirmed in a letter to Carnap that affirmations should be translatable into the language of science and said that his only concern was to deny that they were merely hypothetical (Schlick, letter to Carnap, 05/06/34, referenced in Oberdan, 1996, p.273). The indubitability of affirmations is supposed to come from their grammar (mentioned above) and their demonstrative character. Affirmations are not part of the language of science, because the language of science replaces all demonstratives (such as "now") by the terms which specify what exactly the demonstrative indicates (such as "2:47pm on Tuesday"). Translating affirmations into the language of science involves auxiliary beliefs about when and

where the statement is being made, each of which reintroduces uncertainty. This notion of “translation” is stretched slightly, as the meaning thus preserved is not sameness of intension – it is only extensionally adequate.

Obviously Schlick's epistemology was foundationalist in nature, which is just to say that he considers all truths to derive their justification from these foundational, certain, affirmations. It would be natural to proceed from there to the assumption that his semantics was equally foundationalist - the simplest way for affirmations to provide the justification for other statements is for the meanings of those other statements to be constructed out of some combination of affirmations (e.g. “the sentence 'the square is red and the circle is blue' is true if and only if here now red and there now blue”, for some appropriate “here” and “there”). This would also be consistent with Schlick's well-known adage that the meaning of a statement is the method of its verification (e.g. 1932b, [1979b, p.361]; 1936, [1979b, p.458]), as affirmations on this view are simple statements which are directly verified. Scientific statements would derive their meaning from the affirmations which would provide evidence for them, and affirmations would derive their meaning from the experiences which confirm them (Oberdan, 1996, p.275). Oberdan thinks that this view would have provoked sharp criticism from Carnap. Claims about meaning would be considered meaningless metaphysics under Carnap's thesis of metalogic - they would be an attempt to justify the choice of a language by reference to something outside that language rather than the language's syntactically specifiable features. Carnap's syntactic view was that the only features of language which could be expressed were those regarding the relations between words, and statements about the world were wholly separate from statements about language. That is to say, no single statement could both talk about words and about the world without being a pseudostatement. Oberdan argues that Carnap's commitment to the syntactic view of language goes some way to explaining why he rejected Schlick's epistemic foundationalism - if it's natural to think that epistemic foundationalism and semantic foundationalism go hand-in-hand, then it's equally natural for someone who rejects semantic foundationalism to reject epistemic foundationalism, even if, as Oberdan thinks, the epistemic foundationalism is workable.

Oberdan argues that foundationalism cannot be the correct interpretation of Schlick's semantics because it ignores his early emphasis on implicit definitions (1996, p.276), which we looked at in chapter 1. Terms in Schlick's early semantic system have no content except that which is bestowed upon them by their relation to other terms in the system of science - "for a rigorous science, which engages in a series of inferences, a concept is indeed nothing more than that concerning which certain judgements can be expressed" (AE, p.33). And "[a] system of truths created with the aid of implicit definitions does not at any point rest on the ground of reality," (AE, p.37). This, says Oberdan, is what led Schlick to his "thesis of the incommunicability of contents", that it was impossible to express raw sense data like the greenness of the colour green (see chapters 1, 2, and 3). If Schlick was still committed to implicit definitions in 1934,[25] then it seems his semantics cannot have run parallel to his foundationalist epistemology, because meaning was entirely dependent on the relation of words to each other and could not be expressed by their relation to reality.

Oberdan thinks that by 1934 Schlick had abandoned the thesis of the incommunicability of contents (2013, §9), but not because he thought that language could be grounded in reality. Rather, Schlick had abandoned the notion of content as being anything separate from logical form (Oberdan, 1996, p.279). In his later work (Schlick, 1935a, [1979b]; Schlick, 1935b, [1979b]; Schlick, 1935d, [1979b]; Schlick, 1936, [1979b]), Schlick takes "propositions" to be sentences used in conjunction with certain grammatical rules, and only propositions have the privilege of possessing truth values. Schlick repeatedly emphasised that the choice between alternative linguistic systems is in principle arbitrary - "Which proposition corresponds to which sentence depends upon the grammar that is presupposed, for only the grammar gives meaning to the signs," (Schlick, 1935d, [1979b, p.444]). If the entire meaning of a sentence is determined by grammar, then there is no place

[25] We have argued that he was not, but assume that he was for the sake of following through Oberdan's argument.

in Schlick's semantics for the real world. Oberdan argues that Schlick's emphasis on convention has a lot in common with Carnap's principle of tolerance, even allowing the adoption of languages with radically different logical structures (Oberdan, 1996, pp.280-281).

Schlick's treatment of solipsism in "Meaning and Verification" (1936, [1979b, pp.471-481]) is presented as evidence that sensations such as pain could never meaningfully be spoken about as being entirely private. There, Schlick argues that we can imagine a case in which a person says "ouch" when another person's body is damaged, and our response to this case establishes what we mean by "pain" – either in this case we are feeling another person's pain, because it is that other person whose body has been damaged, or, as he says the solipsist argues, we are feeling *our own* pain in response to the other person's injury. His conclusion is that the solipsist thesis is nothing more than an attempt to get us to adopt a certain rule of grammar:

> ['I can feel only my pain'] becomes a tautology; the word 'can' ... does not denote empirical impossibility, but *logical* impossibility. In other words: it would not be false, it would be *nonsense* (grammatically forbidden) to say 'I can feel somebody else's pain'. A tautology, being the negation of nonsense, is itself devoid of meaning in the sense that it does not assert anything, but merely indicates a rule concerning the use of words.
> - Schlick, 1936, [1979b, p.477], original emphasis

From Schlick's dissolution of solipsism, Oberdan argues that at this time he must have thought words such as "pain" would have to be part of an intersubjective language if they were to be used in meaningful propositions. He says that this would suggest that for Schlick at this time "the privacy of experience is no more than a

contingent happenstance," (Oberdan, 1996, p.282), so that in principle there is no logical reason why one could not experience another's sensations.[26]

If Schlick was broadly in agreement with Neurath and Carnap in terms of language being conventional and intersubjective, why was he not swayed by their arguments that propositions could not be compared to "facts"? Oberdan thinks that the key lies in Schlick's response to Hempel's 1935 paper, "On the Logical Positivists' Theory of Truth", in which Hempel had said that to claim statements express facts is "a typical form of the material mode of speech," (p.54) and therefore liable to produce pseudoproblems. In his response, Schlick wrote "it is easy to express in a purely formal way my opinion that facts and propositions can be compared: words denoting symbols and words denoting other things may occur in the same sentence," (Schlick, 1935a, [1979b, p.402]). Here, "Schlick explicitly confronted the strict dichotomy of the formal and material modes of speech underlying the physicalists' epistemology," (Oberdan, 1996, p.285) so instead of having one class of statements to talk exclusively about the features of the language and another class of statements talking about real objects there would be a class of sentences called "mixed-mode" which would mention both linguistic expressions *and* real objects. In addition to the syntactical rules of grammar a language would have to have application rules - rules which connected linguistic expressions with the world: "the rules of use … do not merely relate, as in logic, to combinations of linguistic signs, but also to the employment of language in life and in the activity of research," (Schlick, 1935b, [1979b, p.408]). Schlick noticed that the rules for certain parts of language, particularly indexicals and demonstratives like "now" and "this", could not be naturally expressed without reference to the situation in the world at

[26] We investigated whether "Meaning and Verification" really shows Schlick to have abandoned the thesis of incommunicability of Content in section 3.3.4, and we concluded that there was no reason to think so. But, again, assume that Oberdan's historical claims are accurate for the sake of argument.

the time of utterance. Whilst Carnap and Neurath had systems of replacing such terms with the objects they represented on the particular occasion of use,[27] coupling the terms with the grammatical rules needed to interpret them appears much more natural. For example, it allows us to interpret sentences written by unknown authors: If we were to find the sentence "I was happy on Tuesday" written on a discarded piece of paper then on Schlick's view we can translate it to "whoever wrote this was happy on Tuesday", but on Carnap's view we wouldn't know what the sentence meant until we identified the author. (We could use metalinguistic rules to infer "this sentence refers to whoever wrote it," but just using rules of the language we wouldn't be able to understand it.)

Oberdan's overall interpretation of Schlick's semantic system looks like this: A statement is a sentence coupled with rules of use, so that the same sentence in two different places may express two different statements if the rules of use are different (or if the rules of use give us a different interpretation depending on location). The rules of use for a sentence can take the form of syntactical rules which purely relate words to each other without reference to reality (e.g. "nothing uniformly coloured is both red and blue"), or mixed-mode rules which tell you how the language relates to the worlds (e.g. 'blue' applies to all and only objects which are blue). And no language can be complete unless there are some rules explaining how the terms of that language relate to the world - in contrast to Carnap's view on the subject, there are no languages composed of nothing but linguistic relations. Protocol statements are those statements which describe experiences in an intersubjective language of science. Affirmations are statements of the form "here now such-and-such", where the such-and-such is sensory information reported in a subjective phenomenal language, but which can be translated into the intersubjective protocol language. Affirmations make use of terms which refer to private sensations, but it is an

[27] See, for example, Neurath's elaborate structure for formalising protocol sentences - "In the universal-slang it is as meaningless to talk of a *personal* protocol as to talk of a *here* or a *now*. In the physicalistic language personal nouns are simply replaced by co-ordinates and coefficients of physical states," (1932, [1959], p.206, original emphasis).

empirical fact that private sensations are coordinated with physical facts (e.g. the sensation of redness with a particular wavelength of light), so there is a relation between affirmations and protocol sentences. On this view, therefore, affirmations are not constructed in a private language, strictly speaking. Affirmations are not protocol statements, even when the term filling the space of "such-and-such" is translated into the language of science, because protocol statements do not contain terms such as "here" and "now", but affirmations can be translated into protocol sentences by replacing the "here" and "now" with the place and time at which the affirmation was recorded. Schlick took affirmations to be certain in virtue of their "here now" character - a certainty which is lost when they are translated into protocol sentences - however (still talking through the lens of Oberdan's interpretation) this certainty was a mistake on Schlick's part because of the intersubjective nature of the "such-and-such" which was being reported – once the thesis of incommunicability of content is dropped it becomes possible to misreport experience in intersubjective language. This is a different kind of uncertainty than the "here" and "now" were meant to eliminate because "here" and "now" both say things about the intersubjective world, but attempting to replace an experience term with "this" says nothing that an interpreter could understand.

I think Oberdan is too quick to claim that Schlick's discussion of solipsism and his acceptance of the separability of a subject from their experiences shows that affirmations must be falsifiable hypotheses in a phenomenal language. Schlick does think that, in theory, any given experience could be experienced by anyone - he states that "primitive experience is absolutely neutral, or, as Wittgenstein has occasionally put it, … immediate data 'have no owner'," (Schlick, 1936, [1979b, p.472]) – but this just means that it doesn't make sense to distinguish between, say, Alice's green and Bob's green. We shouldn't say that when Alice looks at a tree it appears to be Alice-green and when Bob looks at the tree it appears to be Bob-green, because it would be a meaningless tautology to assert "Bob can't see Alice-green" or "the green Alice sees is always Alice-green". The Alice- or Bob- parts are redundant. When Alice perceives a tree as green, this is not meaningfully distinct from when Bob perceives a tree as green. That *does not* imply that the meaning of "green" is a

particular qualitative sensation which is shared by both Bob and Alice. It is entirely consistent with the claim that the meaning of “green” is a formal relation that objects bear to the spectrum of visible light. If anything, it reaffirms the point that we shouldn’t try to say things about content. Furthermore, the picture of language that Oberdan presents as only being Schlick's later position is already present in F&C. On p.310 Schlick talks about how a sentence is meaningless without rules of interpretation, and on p.289 he talks about rules of interpretation as a matter of choice. Indeed, these principles of language are so obvious (i.e. words mean different things in different languages) that it would be weird if Schlick wasn’t committed to them all along. As such, I think it would be a mistake interpret Schlick’s position in a way that demanded content be communicable.

Even if Schlick had abandoned incommunicability of content, this wouldn’t imply that affirmations were supposed to be formed in the intersubjective phenomenal language. In his earlier discussion of affirmations (1935b, [1979b]), Schlick draws a distinction between different uses of the words which describe sensations. Schlick took the sentence "yellow here" to have (at least) two possible interpretations, firstly as an affirmation and secondly as a hypothesis in the language of science. "Yellow here", the affirmation, was to be such that it was rendered meaningless by addition of words like "possibly", but "yellow here", the hypothesis, could accommodate uncertainty in a number of ways, including but not limited to uncertainty over whether "yellow" was the right word to describe the current experience (Schlick, 1935b, [1979b, pp.410-413]). In an affirmation Schlick considers there to be no prospect of making a mistake with regard to which word fills the "such-and-such" position - a person who formed an affirmation with the wrong word is not confused about the content of his experiences, he is just describing them unconventionally. This shows that there was not supposed to be any part of an affirmation which could be interpreted independently of its context of utterance – the meanings of the terms would be fixed by what the speaker was trying to refer to, rather than by an independent standard like interpreter meaning. This means that a better way of describing the structure of affirmations might be "here now this", where according to Oberdan’s phenomenal account “this” will pick

out some content of experience. Oberdan considers this possibility, and we will return to it in section 5.1.3, after considering Uebel's contrasting interpretation of affirmations.

Oberdan's interpretation of Schlick's affirmations resists the usual objection that they cannot be used in support of an intersubjective system of science. However, it ignores the historical evidence that Schlick was focussed on certainty throughout his career and the textual evidence that Schlick repeatedly emphasised certainty in his discussion of affirmations. It also means that Schlick failed to save the correspondence theory of truth - without the certainty of affirmations we are back to a state where all sentences in the system of science are in principle revisable, and so a single sentence does not correspond to a single fact. In the next section we will examine an alternative interpretation, which preserves Schlick's emphasis on epistemology but sacrifices affirmations' foundational role in science.

5.1.2 Uebel's Interpretation – Maintaining Certainty

Like Oberdan, Uebel considers affirmations to be statements, but he does not attempt to ignore Schlick's emphasis on certainty. According to Uebel, Schlick's goal was to "retain the correspondence theory of truth that was part of the picture theory of truth of the *Tractatus*, but replace the original conception of elementary propositions," (2007, p.305). As we saw in chapter 2, elementary propositions take the names of simple objects and combine them in a way which, in some sense, reflects the way the objects are combined in the world. Elementary propositions as they had been conceived would not contain demonstrative or indexical terms – they were formed in the intersubjective language which was learned through elucidation – but affirmations would be of the form "here now X". Affirmations would take on the role of founding the system of science and the certainty of the system would be found in the confrontation of these statements with reality.

Uebel, like Oberdan, draws attention to Schlick's focus on rules of use. Schlick claimed that affirmations are similar to analytic statements in that one knows their truth value as soon as one has properly understood them. But unlike analytic statements, it is not the meanings of the words in affirmations that guarantees their

truth - it is the state of the world. We understand a statement if we understand the conditions under which it is verified, and affirmations are those statements which we verify (or disconfirm) as soon as we understand them, so they are just those statements that describe what we are able to verify in the moment of their formulation. But this does little to elucidate the nature of verification itself, and therefore does not get us any more understanding than we had before.

Uebel thinks that "Facts and Propositions" (Schlick, 1935a, [1979b]) and "On Affirmations" (Schlick, 1935b, [1979b]) can help to shed some light on this issue. In the latter, Schlick wrote:

> Even if it be true that all propositions occurring within the system of science must be viewed as hypotheses, this fact does not entitle us to declare all statements whatever to be hypothetical; there are many propositions which undoubtedly deserve the name 'statements' [as opposed to merely 'sentences'], but are possessed of properties quite different from those of hypotheses, and hence should not be so called. *They do not occur within science itself, and can neither be derived from scientific propositions, nor the latter from them.*
>
> -Schlick, 1935b, [1979b, pp.406-407], emphasis added

This indicates, firstly, that Schlick was more concerned with language in everyday use than within the system of science. This is important, because if affirmations stand outside the system of science then the demand for them to be formed in an intersubjective language is relaxed.[28] The assertion that they "can neither be derived from scientific propositions, nor the latter from them," is a strange one, however, because if there is no inferential relationship one way or the other between affirmations and the system of science then affirmations don't appear

[28] It is not completely eliminated, of course, unless we can address the arguments from Neurath and, later, Wittgenstein, which appear to show private languages to be completely impossible rather than merely bad for certain purposes. We'll return to this issue in section 5.2.3.

to be capable of doing the job for which they were designed. Affirmations could not serve as the foundation of knowledge if all of scientific endeavour was cut off from them, no more than the physical foundations of a building can be placed some way away from the construction site. However, Uebel thinks that even without affirmations being derivable from the system of science (or vice versa), there can still be a relationship between the two. Affirmations can play a psychological role in motivating the acceptance of certain protocol statements, even if their truth does not logically guarantee the truth of said protocol statements.

Schlick's account of verification as something private and only indirectly related to science is borne out in the difference between what he and what Carnap considered to be verifiable in principle. In *Logical Syntax of Language* (1934b, [1937]), Carnap put forward an account on which certain natural laws could be considered part of the linguistic framework of the language – a conventional choice – rather than something to be discovered. He distinguished between two different kinds of inferential rules, L-rules and P-rules, the former being logico-mathematical and the latter physical (*ibid.*, p.180), but wrote "[e]ither L-rules alone, *or L-rules and P-rules*, can be laid down as transformation rules of the physical language," (*ibid.*, p.316, emphasis altered) meaning that, for at least some possible languages, it is possible to interpret P-rules as analytic rather than synthetic facts. This is an example of the kind of conventional choice that Schlick did not think was acceptable, because, for Schlick, grammatical rules, being analytic, cannot say anything about relationships between objects in the world as natural laws should. Carnap's reasoning was that he wanted to be able to assess claims as verifiable in principle, and therefore meaningful, on the basis of rules in the linguistic framework. He argued that if someone postulates that certain objects have a certain property that can only be tested for under faster-than-light conditions, we should be justified in saying that their claim has no place in science (Uebel, 2007, pp.354-355). Schlick, on the other hand, argued that the only things which should be ruled meaningless are those that are *logically* unverifiable, rather than those which are unverifiable as a matter of contingent natural law. He argued that statements about the afterlife are acceptable, with the verification condition "wait until you die" (Schlick, 1936,

[1979b, p.470]). Uebel argues that the difference between the two is coming from the fact that Carnap was interested in language specifically as a framework within which to conduct scientific inquiry, which means that there are different standards at work than apply to our language of everyday use. He was happy to extend his principle of tolerance to languages of the kind Schlick was proposing, and he said that we *could* have languages where the only criterion of meaningfulness was logical verifiability, but he maintained that it was equally possible, and of greater expedience, to rule that only the empirically verifiable was meaningful in science. Schlick, on the other hand, was interested in language as it actually is in the world, and he was therefore only concerned with the restrictions that all languages must share. While Carnap was looking for criteria of scientific significance, Schlick was looking for criteria of meaningfulness in natural language (Uebel, 2007, p.356).

The truth of affirmations was supposed to be guaranteed by the rules of use in a language. Schlick had laid out his most recent account of truth as an alternative to Neurath's in a 1934/35 lecture course:

> A true proposition is one where the rules of language as well as the rules of application have been followed… A proposition is meaningful if it describes a possible state of affairs which accords with the rules of language. A proposition is true if there is something in reality that corresponds to it, if it has been formed correctly according to the rules of application.
> -Schlick, 1934/35, translated and referenced in Uebel, 2007, p.325

The rules of application, as we discussed in section 3.2.1, are a conventional choice. In the making of an affirmation, Schlick says that the speaker selects the rules of application which make it true. This is how Schlick intends to eliminate the error associated with "red" or "green" being intersubjective terms. We may be mistaken that we have ever previously used the word "red" in response to a particular private sensation, but that doesn't matter because *right now* we are making the conventional choice to do so, and therefore "I see red" cannot be a mistake. The word

"red" refers to whatever we are currently seeing - for all intents and purposes it functions as a "this".

Uebel argues that Schlick's conception of truth here is deeply flawed, and indeed it seems that in the last sentence of the above quote where “true” is defined, Schlick is really giving two different definitions which need not necessarily run together. He said a proposition is true if something in reality corresponds to it and then immediately says “if it has been formed correctly according to the rules of interpretation” as though that were just another way of saying that something corresponds to it, but it certainly isn’t obvious that these two things are synonymous. The reason appears to be that the foundations for which Schlick is searching need to be both meaning-theoretical *and* epistemological, and not merely epistemological as Oberdan suggested. If both meaning and knowledge can be founded in the same way, then using words in the correct way and ensuring that they accurately depict the world will come out the same.

On Schlick’s account, if someone is deceived into believing something which they then honestly assert, they are guilty of misusing language in some way. This, Uebel argues, would be highly problematic since intuitively we want to say that competent language-speakers can still make mistakes (2007, p.326). Suppose, for example, I look at a flower and say "the flower is red," but unbeknownst to me I am looking through a red filter, and the flower is actually white. The relevant application rule is that "white" applies to objects that reflect all visible light and "red" applies to objects that reflect only certain wavelengths. My statement is false because I have failed to apply the words in accordance with the rules. If I'm a competent speaker of the language, then it seems like it should be obvious to me whether or not I've applied the rules of the language correctly - I know *a priori*, for example, that any utterance of "the flower is uniformly coloured both red and white" will be a breach of the rules. But in the filter case, it seems that I can't know whether or not I'm following the rules of application properly, which seems inconsistent with my being a competent language user even though there’s no reason to think a competent language user couldn’t be deceived by colour filters.

On the other hand, *something* is obviously wrong with the statement "the flower is red", and the fact that "red" isn't meant to apply to white objects seems like the obvious thing to point to. It does, *prima facie*, seem odd to say that competent speakers of a language won't always know whether or not they're following the rules of the language, but I think this oddness is something we could put up with if it gave us a workable account of the foundations of knowledge. In any case, Uebel and I agree that Schlick's notion of truth is consistent with the picture of affirmations that Uebel is painting. Affirmations are supposed to be those sentences for which we *can* know that the rules of application are followed correctly.

It is one of the effects of these application rules that affirmations cannot be written down. The reason is that the rules of application cannot, in this case, be conveyed along with the statement because they refer directly to part of the world and the world doesn't travel with the statement. In the filter example above we could, and did, lay down the rules of application next to the statement itself. The rules of application in an affirmation, however, involve an ostensive definition. In the language of science an ostensive definition would be replaced by a description of the thing which was picked out, but this is the step that strips the affirmation of its certainty. "[Affirmations'] grammar requires an act, as it were, of direct ostension," (Schlick, 1935b, [1979b, p.413]). It is worth noting that this problem remains when affirmations are spoken, as in Schlick's example of the man reporting his field of vision to a physicist (1935b, [1979b, p.409]). An ostensive reference to a private experience cannot be spoken any more than it can be written down.

Uebel's view is that Schlick no longer thinks we should be asking about how the form of the world is preserved in protocol sentences about the world. Instead, the correspondence which is required for truth will come from application rules that correlate sentences with facts (Uebel, 2007, p.366). Affirmations reach out and touch reality in virtue of these application rules which say "this statement is satisfied by these facts" – these are mixed mode sentences which mention both facts and linguistic expressions, as Oberdan observed. Application rules are fluid in everyday

language, and it is the speaker's intentions that determine them rather than any intersubjective conventions.

So, Uebel's overall picture is one on which Schlick's misguided account of truth led him to look for foundations in statements where truth would be guaranteed by the speaker's own choice of application rules, but which cannot then be connected to the intersubjective system of science. Schlick's account of truth runs together correspondence with correctness of application rules, and speakers get to select their own application rules when they speak, so, when someone makes an utterance with the intention of describing their current experience, that is what happens, and their experience corresponds to that affirmation. "Of affirmations then it can be said that their meaning-theoretical foundations are at the same time epistemological foundations: here what bestows meaning also bestows truth, and vice versa," (Uebel, 2007, p.367). In this way, affirmations function much like a "this". Uttered sincerely, they cannot be mistaken, but we have no way of connecting these utterances to the intersubjective language of science. They can serve a psychological role in motivating the acceptance of certain protocol sentences (*ibid.*, p.332), but that is as strong a link as we can find. We have no reason to think that the application rules associated with an affirmation are the same as those associated with any scientific claim that uses the same words, and so there can be no inferential link from the truth of the affirmation to the truth of the scientific claim.

Uebel's interpretation maintains Schlick's emphasis on certainty, but the only relation affirmations still bear to the system of science is in explaining why we would in fact always class certain groups of protocol statements as being true. Where Neurath and Carnap thought that the source of protocol statements was a question for psychologists, Schlick took it that the answer was affirmations and thought that science had to concern itself with experience rather than mere coherence.

It's true that Schlick did say affirmations bore no deductive relation to the system of science. Uebel's interpretation is in that way true to the text, at least of what Schlick wrote after 1935. However, the spirit of UFE is to attempt to reconcile

certainty with something which founds scientific inquiry. Oberdan's and Uebel's interpretations each preserve one of the two halves Schlick was trying to bring together – on Oberdan's account affirmations fail because their intersubjectivity prevents them from being completely certain, whilst on Uebel's account their certainty prevents them from being intersubjective. Out of the two of them, Uebel's account tracks more closely with the text and is probably a better account of what Schlick intended, but really it might be more accurate to say that Schlick was attempting to bring together two things which intrinsically exclude one another, and therefore whatever position he was attempting to stake out was going to be internally inconsistent. In the next section we will see why neither of the accounts offered has been able to fulfil both criteria, before I go on in section 5.2 to explore why it is understandable that Schlick might have thought affirmations could do the job for which they were intended.

5.1.3 On the Possibility of a Compromise Account

We have seen that Oberdan's account preserved a relationship with the system of science without maintaining certainty, while Uebel's account achieved the certainty objective at the cost of a direct role in science. In this section I will look at whether there is room for an account between the two. Is it the case that Oberdan's account is as close as we can get to certainty without sacrificing our link to the system of science? And, in the same vein, is Uebel's account as close as we can get to intersubjectivity without sacrificing certainty?

Taking Oberdan's account as a starting point, our problem is that we cannot seem to establish certainty for intersubjectively valid statements. We started by acknowledging that we could never be 100% secure in a protocol statement of Neurath's construction, which looked like this: "Otto's protocol at 3.15: [at 3.14 Otto wrote: (at 3.13 there was a table in the room perceived by Otto)]". There are all sorts of ways that statement can go wrong, e.g. if Otto is wrong about the time or place where he saw the table. So, we got rid of specific terms and replaced them with demonstratives and indexicals – these would be lost in the translation back into a useable statement for the system of science, but would help us to secure our

observation statements at the moment in which they occurred. We got rid of the term denoting the place of utterance and replaced it with “here,” removed the term denoting the time of utterance and replaced it with “now,” and formed the statement first-personally instead of declaring ourselves to be Otto. We also simplified the table down to component sense data. That left us with something like “here, now, brown”. That’s still intersubjective, because “here” refers to a point, wherever it might be, that can be identified by any interpreter – space is shared between speaker and listener. Likewise, time remains the same for the speaker and the listener. “Brown” might refer to an external property or, on Oberdan’s account, an internal property properly correlated with the external world, but in either case uncertainty remains in the formulation because of the familiar problems of illusion, hallucination, and forgetting what the word “brown” means. The logical conclusion of this line of thought is to replace "brown" with "this" and leave no room for error whatsoever.

Oberdan rejects affirmations so construed as being able to play any part in a viable semantic system for two reasons. The first is that "The element of certainty thus isolated at the foundations of knowledge is purely subjective," as the indubitability obtains "only for the speaker at the time of utterance". I think we can set this objection to one side, as it was a feature of affirmations which Schlick made explicit from the beginning – affirmations are only supposed to be certain for the person who forms them. Additionally, the epistemic properties of any theory, whether it’s formed in an intersubjective language or not, will always be relative to a subject – that’s just part of the nature of epistemic properties. Even *a priori* truths are only known to people who understand the reasoning. Oberdan’s second point, mirroring the original objections of Neurath and Hempel, is that the expression "here now this", being completely uninformative on the surface, cannot possibly be used for communication. If “this” is directed towards an external object then the speaker can’t know for sure what they’re indicating, and if anyone is going to know what an affirmation means then the person uttering it should probably know. If, on the other hand, "this" is directed towards a sensation, which is the only way that it's going to be able to give us the certainty we want, then it's no longer possible for a

listener to interpret the statement. This is a problem which arises for "this" in a way that it doesn't for "here" and "now", since experience is relative to the subject but speakers and listeners both presumably occupy the same spatiotemporal universe. In other words, it is unclear how affirmations could relate to the intersubjective system of science if this is what they say. What this means is that if we start from the assumption that affirmations are formed in an intersubjective language and try to establish certainty, we are doomed to failure.

Starting from the Uebel direction and working towards intersubjectivity, another problem arises: inference runs both ways. Suppose we have absolutely certain affirmations and we want to relate them to the system of science such that the affirmations stand in a support relation to the rest of the system. If P supports Q, for some suitable support relation, then ¬Q supports ¬P. This is the principle of contraposition in formal logic. The problem is that if the truth of an affirmation supports the truth of a theory, then it is also the case that the falsity of the theory supports the falsity of the affirmation. But if some circumstance is able to support the claim that an affirmation is false then affirmations are not epistemically untouchable, which violates our initial assumption. Therefore, if affirmations are epistemically secure, then we cannot put them into a support relation with the system of science.

So, it looks as though any attempt to make an intersubjective statement epistemically secure will strip it of its meaning, and any attempt to connect epistemically secure propositions to the intersubjective system will introduce a new source of risk to those propositions. At this point it looks as though any interpretation is going to fall short of one of Schlick's objectives. The question now becomes, why didn't Schlick see this as a problem, even when everyone around him was trying to point it out? In the next section I will explore whether the account of language that was emerging from developments in Wittgenstein's thought might have influenced Schlick's position on this matter. I will argue that he had hoped to use a Wittgensteinian criterial relation to epistemically connect affirmations to the system of science without introducing a deductive relation, perhaps allowing us to

justify beliefs in the system of science without risking the certainty of the affirmation if revisions to the system become necessary. I will ultimately argue that the criterial relation is underexplained, and therefore doesn't really allow us to bridge the gap, but also that it probably does explain what Schlick was trying to do.

5.2 Reconciling the Aims

In this section I will attempt to explain how Schlick thought it would be possible to preserve the two crucial elements of the foundations that he was looking for – certainty and relevance to the system of science. We will start by examining the text of UFE for as much information on the nature of affirmations as possible, and will build up a picture of what they must be from that. In section 5.2.2 I will draw together the necessary features of affirmations with what we know about Schlick's philosophy of language, that is, with the features shared by all propositions. At this point, private language arguments become a concern, and I will take section 5.2.3 to discuss the ongoing influence Schlick was under from Wittgenstein. We will see that affirmations can be construed as a special case of Wittgenstein's account of reasons in the "Diktat für Schlick", and this introduces the Wittgensteinian notion of criteria into the picture as something separate from truth conditions but nevertheless contributory towards meaning. In section 5.2.4 I will sum up the view of affirmations which I believe Schlick was attempting to elucidate, which is as the criteria for the justified acceptance of a protocol statement, but we will see that Wittgenstein's criteria don't really do anything to solve the central problem of getting epistemically secure propositions to play a role in a system of revisable scientific claims. In section 5.2.5 I will address the intersubjective-infallible dichotomy and argue that the only way to reconcile the two is with an account on which not only affirmations but also the system of science itself is interpreted in a private language, although this approach comes with its own major drawbacks.

5.2.1 "On the Foundations of Knowledge"

Early on in UFE, Schlick writes:

> It makes no sense to speak of uncertain facts – it is only statements, only our knowledge that can be uncertain; and if it is therefore possible to reproduce the raw facts quite purely in 'protocol propositions', the latter seem to be the absolutely indubitable starting-points of all knowledge.
> - UFE, p.370

In a moment, Schlick will go on to consider the ways in which uncertainty may creep into our system, but it is already clear that he does not believe it has to appear at the first moment. If we have some procedure by which a proposition may be made to reliably mirror a fact in the world, that proposition can form the foundation on which we build our knowledge of the world. I would also point out at this stage that, whilst Schlick brings up what protocol sentences were originally supposed to be so that he can show why they don't work, he nevertheless implies that the conditions for their success would be *reproducing* facts, not merely establishing some kind of arbitrary correspondence with them. This suggests that there is still more to meaning at this stage than we saw in Uebel's interpretation – statements still picture the world in a certain configuration and show something of the structure that they say obtains.

Immediately after the above passage, Schlick writes:

> They are indeed left behind, the moment we go over to propositions which are really useable in life or science (such a transition appears to be that from 'singular' to 'general' statements),
> - UFE pp.370-371

"They" here appears to refer to protocol propositions, but at this point in the essay that does not necessarily refer to sentences of Neurath's complex construction. Schlick is instead talking about *whatever* it is that we want to fulfil the foundational role in the system of science. From this second passage it seems that he thinks that we can have absolutely certain statements about restricted "singular" events or things (perhaps "the raven is black"), but also that the certainty is

unavoidably lost when we attempt to move from these to some general hypothesis ("all ravens are black"). The latter are clearly inferred from (some collection of) the former, but the inference relation between them is not logical implication and they don't therefore carry the same degree of certainty. Nevertheless, the general statements are the ones from which we are able to make future predictions ("the next raven I see will be black"), and they are therefore the real subject matter of scientific inquiry.

Next, Schlick explicitly rejects the idea that protocol propositions, rightly construed, should be facts in the world.

> [I]t was proposed to regard as 'protocol propositions' certain spoken, written or printed sentences, i.e., certain sign-complexes consisting of sounds, writing fluid or printer's ink, which if rendered out of customary abbreviations into a fully articulate language would signify somewhat as follows: 'Mr So-and-so, at such-and-such a time and place, is observing this or that'. (This view was particularly upheld by Otto Neurath.)
> - UFE p.372

> I believe, in fact, that the view arrived at by these considerations about protocol propositions is untenable. They lead towards a peculiar relativism, which seems to be a necessary consequence of the view which regards protocol propositions as *empirical* facts, on which, as time goes on, the edifice of knowledge is raised.
>
> For as soon as one asks about the certainty with which one may maintain the truth of protocol propositions regarded in this fashion, one has to admit that it is exposed to all manner of doubt.
> - UFE p.373, emphasis added

Schlick thus appears to say that we should not look to physical facts – utterances and written sentences – to ground the system of science. True protocol propositions (i.e. affirmations) are not to be found in the world. However, recalling

the results of chapters 2 and 3, we know that expression is a relation between facts. If affirmations *say* anything, they *must* be found somewhere in the world. In fact, I think that the crucial distinction Schlick is making in this passage is between physical facts and phenomenal facts. The two are, of course, expressed in mutually translatable languages, but Schlick thinks that the process of science involves deriving the physical from the phenomenal rather than the other way around. Schlick rejects the thought that we might find a sentence written in a book and declare it to be beyond dispute. Instead, the facts which constitute indubitable propositions are to be formed internally by us, and will hold their indubitability only at the moment of formation.

Backing up this account of affirmations as intrinsically distinct from the propositions we might find in a book, later on Schlick writes:

> Whether and to what extent we declare a statement to be corrigible
> or capable of annulment depends entirely *on its origin,*
> - UFE, p.378, original emphasis

This is in the context of rejecting the economy principle – the idea that we can have a unique criterion of truth in a coherence system by always adopting as true those statements which require the smallest change in the system to maintain. However, Schlick's comments here have broader implications. Schlick thinks that we cannot take statements of a particular form as foundational to the system – statements of the form "at 3 o'clock Otto saw a table" have an identical structure when uttered sincerely as they do when uttered as a joke (UFE, p.378) but we nevertheless ascribe more weight to one than the other, and that tells us that the form of the statement alone is not what matters in our decision about which statements to take as foundational and which to reject. This is in direct conflict with the syntactical account that Carnap was maintaining at the time. What matters, Schlick says, is where the statements come from. We don't pay any attention to statements uttered as a joke, we pay some attention to statements formed from distant memory, and we pay most attention of all to statements formed in response to present instants of experience. Interpreted more broadly, Schlick is arguing that

context of utterance plays a key role in not only the determination of the meaning of an utterance, but in the question of what it would take to verify or falsify the utterance. This will be a point of great importance when we turn to the Wittgensteinian notion of "criteria" to shed light on Schlick's account, as context plays a key role in determining what the criteria for a certain claim will be.

> For in this ordering of propositions according to their origin, which I undertake in order to judge their certainty, an exceptional position is immediately taken by those which I advance myself. And of these, a secondary position is again occupied by those lying in the past, since we believe that their certainty can be impaired by 'deceptions of the memory' – and the more so, in general, the further back in time they lie.
> - UFE, p.379

Schlick had always advocated a more solipsistic account of knowledge than Carnap or Neurath, and here it was to play a crucial role. Schlick goes on to point out that if everyone else in the world made assertions that were different to our own observations, with only enough agreement to establish a common language between ourselves and them, then we would always take our own observations to outweigh theirs. If everything else about the world was exactly as it is, but the sky appeared to us to be green, then all the textbooks in the world could not convince us that the sky appeared to be blue.

> It is theoretically conceivable that the statements made by everyone else about the world should be in no way confirmed by my own observations. It might be the case that all the books I read and all the pronouncements I hear are in perfect agreement among themselves and never contradict one another, but that they are utterly irreconcilable with a large part of my own observation propositions. (In this case, the problem of language-learning and its use for communication would create certain difficulties, but they are soluble by means of certain assumptions as to

> whereabouts alone the contradictions are to appear.) According to the theory under criticism [Neurath's (alleged) coherence theory of truth], in such a case I would simply have to sacrifice my own 'protocol propositions', since they are certainly at odds with the overwhelming mass of the others, which do harmonise together, and which there can be no possible expectation of correcting by reference to my own limited and fragmentary experience.
>
> But what would really happen in the case supposed? Well, I would not give up my own observation propositions under any circumstances, for I find, rather, that I can only adopt a system of knowledge which they fit into without mutilation. And such a system I should always be able to construct. I need only regard other people as dreaming fools, whose madness has uncommon method in it – or to put it more concretely – I would say that the others are actually living in a different world from mine, which has only just so much in common with the latter as to permit communication in the same language. In any event, and whatever the world-picture I construct, I would always test its truth only by my own experience; this support I would never allow to be taken from me, my own observation propositions would always be the final criterion. I would proclaim, as it were: 'What I see, I see!'.
>
> - UFE, pp.379-380

This appears at first glance to be an empirical claim rather than a philosophical one ("what would *really* happen in the case supposed?" instead of "what *must* happen…") but the character of Schlick's claim stands out more clearly when he says: "to put it more concretely – I would say that the others are actually living in a different world from mine," and "whatever world-picture *I construct*, I would always test its truth only by my own experience," (emphasis added). Schlick had previously argued, both throughout "Positivism and Realism" and the third lecture of "Form and Content" (pp.335-336), that we could not meaningfully distinguish between the internal and external worlds. All that exists is that which is

experienced. This is important background to his development of affirmations, and we will return to it in the next section.

Schlick has not, at this point, laid out what the foundational statements are supposed to look like, or even used the term "*Konstatierungen*", but he does now explicitly turn to the two questions that we have seen previous accounts having so much difficulty in answering consistently: "What actually lies behind the statement that they are 'absolutely certain'? And in what sense can they be designated as the ultimate ground of all knowledge?" (UFE, p.380). In other words, what grounds their certainty, and how do they relate to the system of science?

In his answer to the second question (which he addresses first) Schlick says:

> If we suppose that I at once take note of every observation – it makes no difference in principle whether I do so on paper or in memory – and now start out from thence to construct science, I would have before me genuine 'protocol propositions',
>
> …
>
> It is clear what role is played in [the actual procedure of science] by assertions of the 'immediately perceived'. They are not identical with statements written or remembered, i.e., with what could properly be called 'protocol propositions', but are the *occasion* for framing them.
> - UFE, pp.380-381, original emphasis

This passage creates confusion, because now it looks as though affirmations are not supposed to be statements at all, but events of some kind. The reason for this, as we shall see, lies in how Schlick intends to establish an affirmation's certainty. Affirmations necessarily involve a demonstrative term, which means that they can only be properly interpreted at the moment of utterance. The utterance of an affirmation is not the same as the utterance of a statement in the system of science because a scientific statement says the same thing on each occasion, whereas the meaning of an affirmation is linked directly to the context of utterance,

including the moment in time at which it occurs. It's not even possible to give an affirmation's meaning as a function of context, because it appears as though Uebel is right to interpret Schlick as demanding that the rules of interpretation be fixed by the speaker at the moment of utterance, which makes the rules themselves context-dependent and that precludes the possibility of saying in advance of knowing the context how we should interpret a statement in that context. This means that while affirmations are, strictly speaking, statements (certainly as we understand the term today), they also have the spatiotemporal features of events.

When it comes to the question of establishing the certainty of affirmations, we've seen that Schlick draws an analogy between them and analytic statements. An analytic statement is such that as soon as you have understood the statement, you know, in virtue of its form, that it is true. Similarly, affirmations are supposed to be such that as soon as you have understood them, you know whether they are true or false. The reason that this seems like it might work stems from the affirmation's demonstrative character. As Oberdan rightly pointed out, Schlick emphasised the need for mixed-mode sentences, which means that terms about language and terms about the world can occur alongside one another in the same sentence. In an affirmation, there is always a demonstrative "here, now", picking out some feature of the world, and when you understand the affirmation, you understand what the "here, now" is picking out – specifically, this will be some feature of present experience. Whatever is happening here, now, it is accessible to you – you are experiencing it because whatever you are experiencing is here and now. That allows you to compare it with the other half of the proposition. So, you know whether or not "here, now, black marks on white paper" is true because if you understand "here, now" you understand that it refers to these black marks on white paper, and therefore the sentence has the same structure as the analytic statement "where black marks occur on a white background, there is a white background with black marks on it". Now we must again face the issue which has been highlighted repeatedly in the discussion of Uebel's and Oberdan's accounts – does an affirmation's "here, now," refer to an internal feature of experience or an external one? The first option causes problems for intersubjectivity, but the latter foils the

epistemic objective. Schlick's answer, as hinted at above, is going to be that this isn't a meaningful question.

5.2.2 Certainty

It seems that in order for affirmations to have the epistemic character of analyticity, the "here, now," cannot reference the intersubjective world. If the claim, "here, now, blue" is interpreted as "somewhere in my presence there is something blue" then there is no guarantee of its truth, given the possibility of being deceived about things which appear to be in our presence. This seems to leave us back where we started. Either affirmations refer to something in the real world, in which case they're uncertain, or they refer to private sensations in which case they cannot found the system of science. How did Schlick think he would be able to bridge this gap?

Schlick cannot have thought that affirmations would be formed in a private protocol language along the lines that Oberdan suggested, because Schlick denied that anything which could rightly be called "language" would be able to express content. "Black marks on a white background" doesn't denote qualia, because nothing does. Rather, it must denote a formal relationship between some marks and their background. In order for experience to tell us how the world is, experience must represent the form of the world in its own material, whatever the material of experience may be said to be. This was the result of our discussion of Form and Content in chapter 3. Our experience is accurate (i.e. not an illusion or a hallucination) if and only if the facts in the world share the logical form of the experience. Importantly, though, this implies that there is *some* matter of fact about the logical form of our experience, and that form is the part that we want to express when we make statements about our experience.

Schlick wants to make these statements about the facts of experience in such a way as to guarantee their truth – this is what he stated as the goal of true protocol propositions in the opening pages of UFE: "if it is therefore possible to reproduce the raw facts quite purely in 'protocol propositions', the latter seem to be the absolutely indubitable starting-points of all knowledge," (UFE, p.370). It's not

possible to reproduce the facts of the external world infallibly – hallucinations and illusions are always possible – but we might be able to reproduce the facts about the way things appear to be. Thus, when we say "here, now, black marks on a white background" we are saying something about that matter of fact in the world which is our current experience. Affirmations are statements about experience, which makes them statements about something in the world (and this in turn means that they must be related to the system of science, which should aim to describe the world in its totality), but they are not statements about the stuff that we are experiencing. "Here, now, black marks on white background," can still be objectively true, even if the thing experiencing the writing is a brain in a vat.

We might now worry that if experience is a matter of fact and affirmations are statements about that fact, affirmations must open themselves up to error because it is always possible that the fact does not obtain. This is the line taken by Neurath in his response to Schlick, where he again points out that it is a feature of any true proposition that it be subject to disconfirmation (Neurath, 1934, [1983, pp.102-105]). This line of reasoning risks failing to make a distinction between two different ways in which we can say something is "possibly false". Any meaningful proposition of the form "here, now, φ" is possibly false in the sense that there must be possible worlds in which φ does not obtain at the location picked out by "here, now". But it does not follow that it is *epistemically* possible that "here, now, φ" is false. It is possible to see this kind of confusion as being behind Schlick and Neurath's persistent disagreement, since Neurath was clearly interested in a logical property of propositions while Schlick was trying to argue that there are synthetic truths that could nevertheless be known. Clearly, if synthetic knowledge is at all possible then it is possible to know something which is possibly false, because if it wasn't possibly false then the knowledge wouldn't be synthetic. It would be disingenuous to reply to this, "but how could you know it if it might be false?", because that only carries rhetorical weight when the "might" comes with an implicit "for all you know", and *ex hypothesi* it is *not* the case that the supposed piece of knowledge might be false for all we know.

If I know P, it is not epistemically possible that P is false. It seems as though I must be able to know that "here, now, φ" is true if I am currently experiencing φ (which is a stronger condition than *if I am currently having an experience as if of φ*) and I understand the "here, now," to be pointing to that which I am experiencing. "φ" here is to describe logical form which, as a matter of fact, is instantiated by the experience of the subject and which may or may not also be instantiated by the external world. The "here, now," points to a structure of which we are directly aware, because what it points to is that-of-which-we-are-aware. If the structure pointed to by the "here, now," is the same as the structure of the second half of the sentence then we know that the affirmation is true in the same way that we know the truth value of an analytic proposition.

Now is the appropriate point to introduce concerns about private languages. There are a couple of different worries that we might have about the grammar of φ in the above examples. Firstly, if we're assigning a meaning to φ that makes it true that we're currently experiencing whatever φ denotes, then we're *not* assigning a meaning to φ based on what other people take φ to mean, or on what we ourselves might have taken φ to mean at another time. The immediate upshot of this is that φ cannot reliably be used for communication. When meaning depends upon features of the world not accessible to the listener, the meaning of a sentence is inaccessible to the listener and the notion of communication breaks down, and a language which can't be used for communication is not much of a language. Secondly, if φ means something like "whatever I'm currently experiencing" then it's not at all clear that it's *logically* possible for "here now φ" to be false. As Wittgenstein put it in the *Philosophical Investigations* (henceforth "PI"), "One would like to say: whatever is going to seem correct to me is correct. And that only means that here we can't talk about 'correct," (1953, [2009, §258]). That is, when the meaning of a term is *whatever it needs to mean to make true the statement in which it occurs*, it doesn't make sense to ask whether the statement in which it occurs is true, which is another way of saying that it isn't really a statement and the term in question is meaningless. The first worry, if it applies to affirmations, would make it obviously difficult to link them to an intersubjective system of science, but it at least allows that they could be

false in the logical sense just highlighted. The second worry is that affirmations formed in a private language aren't really formed in any language at all, for exactly the reasons Neurath is pointing to – if they can't logically be false then they can't be meaningful.

Wittgenstein's PI was, of course, much later than Schlick's UFE, but Schlick should have been sensitive to similar concerns from the earlier "Diktat für Schlick" ("Dictation for Schlick"), which Wittgenstein dictated to Waismann (presumably for Schlick) shortly after the Form and Content lectures of 1932.[29] Schlick's claim "immediate data 'have no owner'," (1936, p.472) referenced above in our discussion of Oberdan, appears here by Wittgenstein (Wittgenstein & Waismann, 2003, p.27). In investigating the meaning of the word "understanding", Wittgenstein asks whether we can make sense of the idea that a proposition is accompanied by a thought, and says: "For us, there is nothing essentially private about thinking," (*ibid*, p.27). While he does not explicitly say at this point that we can't make sense of the idea of a private language, the thrust of the piece is that understanding a proposition is a public matter. The same considerations are the topic of the first chapter of *Philosophical Grammar* (1974, §1-§13). Wittgenstein endorses a strongly-behaviourist view of mental processes – "Every so-called inner process is replaceable for us by an outer one, a memory image by a painted picture, conviction by a gesture of conviction, etc.," (2003, p.27) – which seems like it would rule out an appeal to private sensations as either the meaning of a proposition or the foundations of the system of science. There is, however, a strong similarity between

[29] There is some recent debate about the precise dating of "Diktat für Schlick", with Manninen (undated) suggesting it could have been as late as 1935, after UFE was published, and even that it might have been dictated by Waismann rather than Wittgenstein. I think that the parallels discussed above between Schlick's writing and the *Diktat* are too close for Schlick not to have been inspired by the new Wittgensteinian ideas, but we don't need to assume that Schlick had a copy of the work in his hands while he was writing. Whatever the date of the *Diktat* itself, Schlick was in such frequent contact with Wittgenstein (and Waismann) over this period that we may assume ideas were shared earlier than they were put to paper. We can also find evidence of Wittgenstein's position in *Philosophical Grammar*, which is generally dated to before 1934 (Rhees, 1969, [1974 pp.487-488]), and where appropriate I will draw attention to this as we proceed.

Schlick's assertions about the grammar of affirmations and Wittgenstein's discussion of expectation and reason for action. Wittgenstein writes:

> How does an agent himself know what he expects? Does he observe his own behaviour and conjecture from it that he really expects Mr N for dinner? If we say that he must surely know whether he expects this person, then this proposition resembles the proposition that he must surely know the reason for his own act. Someone asks me: 'Why are you turning out the light in your room?' I say: 'Because I want to go to sleep'. He asks: 'Are you sure?' And I reply: "I must surely know why I am doing it'. This certainty indicates that specifying a reason is the criterion for having this reason.
>
> - Wittgenstein & Waismann, 2003, p.31

What Wittgenstein is saying here is that it is a feature of the grammar of sincerely-asserted reasons that it doesn't make sense to question whether the speaker has correctly identified their own reasons for doing something.[30] Wittgenstein's notion of "criterion" is one of justification for assertions. Why do we use the term "X" for this object? Because it meets the criteria of X. Why do we say that it meets the criteria of X? This latter question is simply misconceived (Wellman, 1962). This is the same feature of grammar that Schlick says accompanies an affirmation – that there is some use of "here, now, blue," for which it just doesn't make sense to ask, "are you sure?" – suggesting that Schlick had a Wittgensteinian conception of language in mind when he was describing them. Wittgenstein's idea of a criterial relationship seems like it might go some way towards explaining what Schlick thought the relationship of affirmations to the system of science might be.

---

[30] The situation is complicated by the possibility of someone lying about their reasons – Wittgenstein puts this down to language not behaving according to strict rules, and begins making use of the concept of family resemblance to deal with the various different ways people might have reasons for their actions (pp.35-37), although the term "family resemblance" doesn't appear until his later works. We will have more to say about lying below.

5.2.3 Wittgensteinian Criteria

What are criteria, as Wittgenstein conceives of them? They are not mere truth conditions, but they are something to do with truth and evidence. Hacker (1986, pp.308-309) writes that in Wittgenstein's earliest use of the term "criterion", which was in his 1932 Cambridge lectures, he said that we can determine one proposition to be a criterion of another proposition as a matter of grammatical convention. These grammatical conventions will, at least in part, fix the meaning of the terms thus employed, but they are not *equivalent* to the meaning. For example, "S says that they have a toothache" can be a criterion for "S has a toothache", in the sense that the first proposition is part of the verification of the second, but it would be wrong to say that "S has a toothache" just means that S says they have a toothache, since it is obviously possible for S to have a toothache without vocalising the fact (though we would not say it is possible for them to have a toothache without there being any physical signs, so this just means that "S is holding his inflamed jaw and groaning" is an alternative criterion – it is not saying that S might have a toothache without satisfying *any* criteria). However, Hacker points out, a couple of years later, in the Blue Book, Wittgenstein distinguished between criteria and symptoms. That a person says they have a toothache is now a symptom, not a criterion, of their toothache – the criteria of a person having a toothache will instead involve certain physical states such as inflammation in the person's jaw, and this is because we think that the inflammation is necessarily related to the toothache whereas the complaining is only contingently related. So, in the later work, criteria appear to take on a more focussed role as determiners of meaning. The distinction between criteria and symptoms is that criteria are defined by grammatical convention whereas symptoms are discovered regularities in the world. We discover that one thing is a symptom of another by having an independent grasp of what each thing is and then finding them together in the world, whereas criteria cannot be separated from that for which they are criteria – we can't identify a toothache independently of identifying a person with an inflamed jaw. However, Wittgenstein also points out that in actual usage the distinction is not clear-cut and stable over time. There is not always agreement between doctors over whether certain symptoms are associated

with a certain disease by empirical regularity or by definition, and over time things that were thought to merely correlate with a disease may be recategorized as *what it is to have* that disease. The idea that what counts as a symptom and what counts as a criterion will change over time is retained all the way into PI (§354). Of course, "change" doesn't mean the same thing as "overlap" – we might say that at any one time something can be either a symptom or a criterion for something else, but never both at once, and this is obviously the kind of picture Schlick has had of scientific concepts since AE (see our discussion of "all bodies are heavy" on pages 17-18 above). In the public language, the meanings of our words have to be relatively stable and so this shift between criteria and symptoms describes a long-term feature of the language, rather than an utterance-by-utterance choice on the part of the speaker.

The link between *P* and the criteria for *P* is stronger than an evidential one. An evidential relationship has a certain kind of imperfection: it leaves open the possibility that there may be stronger evidence available. Criteria, on the other hand, are supposed to be so tightly bound up with that for which they are criteria that it doesn't make sense to question the latter if the former are confirmed. Hacker offers the example of a person falling, breaking his leg, and lying on the ground groaning, as related to the claim that this person is in pain (1986, p.312). It would be wrong, on Wittgenstein's account, to say we have good "evidence" for the claim that they are in pain, because that would imply that there might be better evidence and that we can make sense of the idea of doubting the claim. Rather, we should say that the criteria for this person being in pain have been met. It is inconceivable that we should come to know that they are in pain in any way other than by observing what injuries have befallen them and the way that they have responded.

It is important to see that, on this criterial account, "S is in pain" does not mean the same thing as "S is lying on the ground clutching a broken leg." One reason for this is that S could have broken their arm instead, but the existence of multiple different criteria for the same thing is only part of it. The real issue is that, while we can *see* that S is lying on the ground clutching a broken leg, pain is not the kind of

thing that can be seen. That is just to say, pain and writhing are not the same thing. This issue comes out more clearly when Wittgenstein discusses the criteria for someone having an ability to do something (e.g. PI, §148-§149). If we observe someone making moves in a game of chess that are allowed by the rules and that move towards the objective of capturing their opponent's king, then we will say that they meet the criteria for knowing how to play chess. But they don't stop knowing how to play chess when the game is over. They still know how to play chess, even when they are not currently fulfilling the criteria by making moves, and therefore we must say that the state of *knowing how to play chess* is different than the state of *making permitted moves in a game of chess*.

Despite the separation between criteria and that-for-which-they-are-criteria, Hacker argues that it would be a misconception to think that it is a matter of inference to move between the two. We do not observe the criteria and then deduce what they imply about the state of the subject – observing that the criteria of *P* are fulfilled is the same thing as judging that *P* is true. As stated above, there is nothing *more* to S being in pain than their injury and response to it – their pain is manifest in their bodily state and actions – it's just that pain isn't the kind of thing that we observe (or, as Hacker points out, fail to observe), whereas injuries and behaviour are.

We must also note that criteria are highly context dependent, because what it takes to verify a claim will depend a lot on the situation. For example, whether or not screaming is a criterion for pain depends on whether or not the subject is on a roller coaster. This context-dependence has some interesting epistemological consequences in Hacker's account of criteria (*ibid.*, pp.317-318). The link between *P* and the criteria for *P* is one that preserves doxastic states – if you are certain that the criteria are satisfied, then you are certain of *P*. On the other hand, criteria seem to be defeasible. Clearly, we might be tricked into thinking that S is in pain, only for them to jump up from the ground and say, "fooled you!" Hacker attempts to reconcile these by arguing that for something to be *defeasible* is not the same as for it to be *defeated*: "the claim that the satisfaction of the criteria *p* confers certainty, proves,

or is decisive for the truth of *q* is compatible with the claim that the criteria are defeasible," (*ibid.*). When S has jumped off the ground, the context is one in which S's having been writhing around a moment ago is insufficient to verify pain, and so the criteria of S being in pain are not met. In the context where S doesn't get up, however, the writhing around *is* a criterion of S being in pain. The criterion is still defeasible, and we know this because we just described a situation in which it was defeated, but, in this context, it has not been defeated and so it can confer certainty.

We saw before that affirmations are supposed to at the very least provide motivating support for protocol sentences. If asked, "What makes you think that you saw blue on Thursday at 2pm?" then we may respond "Because here, now, blue." This gives us some explanation of the relationship between affirmations and the system of science, but it is important to remember that criteria are not what establish affirmations as *true*. The two parts of the relationship at this point are *that subject S asserts affirmation A* and *that we adopt protocol sentence P for S*. The former is a criterion of the latter. It is the observable fact that allows us to know what S's experience is, and the relationship here is one of justifying the adoption of the protocol sentence, not one of knowing that the truth conditions for either of A or P are satisfied.

Criteria being a matter of publicly-observable grounds for acceptance, they are not the right tool for considering why affirmations are supposed to be infallible. The infallibility of affirmations has always been grounded in their first-person character, and Uebel is right to interpret Schlick as demanding that their rules of application be selected so as to make them true – this is also what we saw to be the case in the last section. We can still interpret affirmations as public statements and say that the criteria for their acceptance is that they are asserted, which does appear to be Schlick's position, but *what they mean* is determined by the speaker. Criteria play the role of getting affirmations to relate to the system of science despite their having been formed in a non-intersubjective language, but they don't give us the truth conditions or infallibility of the affirmation in the idiolect of the speaker.

5.2.4 Summing Up Schlick's Affirmations

Making affirmations work requires a lot of philosophy that is almost unique to Schlick. They emerged out of his commitment to empiricism and his search for secure foundations for science. They therefore needed to be at once connected to the intimately first-personal realm of experience, and the intersubjective world of communication of facts and records of data. They are bound up with both an early-Wittgensteinian conception of language as strict rules-based calculus, and with the transition to a looser account of language as a messy feature of life.

For the certainty that Schlick requires as a secure epistemic foundation, an affirmation must in some sense point to that which it is about. Propositions are combinations of sentences and rules of interpretation. In the intersubjective language, the rules of interpretation are a feature of the linguistic community, so that it makes sense to criticise a person's choice of words to describe something and we can teach people how to use language properly. The rules of interpretation for the affirmation, however, must be selected by the speaker at the moment of utterance. These rules, because they contain an act of direct ostension, cannot be transmitted along with the sentence that is uttered, and so the affirmation as a proposition is essentially incommunicable. It cannot be written down, and therefore strictly speaking it can only be properly interpreted by the speaker. It is a recurring claim in Schlick's philosophy that languages like this are possible.

In positing these incommunicable propositions, however, Schlick was not falling afoul of private language concerns. The rules of interpretation for an affirmation are not, and could not be, that they mean whatever is currently true. Neurath was right to argue that it is a feature of a meaningful proposition that they can be true or false, and affirmations do not violate this rule, *provided that we understand the 'can' to be of a logical, not an epistemic, kind*. Affirmations do not say "here, now, this," and Oberdan was quite right to say that if they did, they would be meaningless. Affirmations say, for example, "here, now, blue," and come along with a rule that says "'blue' refers to this," and the distinction is important because the subject's adoption of that rule implies that, if something other than *this* had been present, 'blue' would not have been the appropriate word to use (whereas "this" is

always an equally-appropriate, and equally-uninformative, word to use, no matter what is in front of you). The subject's idiolect at any given moment comes with a whole host of different words to use in different situations, and there are other sentences that the subject could have uttered in place of "here, now, blue," such as "here, now, red," that would have been a lie. The infallibility that Schlick was emphasising for affirmations is of an epistemic kind, not a logical kind, which means that sentences with the form of affirmations have the possibility of being false, but that nevertheless some of them will be known to be true.

The role that these infallible statements were to play in scientific inquiry was not one of deductively implying scientific claims – that was made impossible by the fact that they were not formed in the intersubjective scientific language. Schlick was, as Uebel identified, following Wittgenstein into a conception of language that was less rigid than previous accounts, and part of that relaxation was the introduction of the notion of *criteria* to the picture, as something related to, but importantly separate from, truth conditions. That a person is prepared to assert "here now blue" at a particular moment, selecting their rules of application so as to make the statement a true affirmation, is what fixes the meaning of "blue" for them at the time of utterance. In contrast, in the intersubjective language of science "blue" is defined by some kind of agreement across the population. Important at this point is Schlick's repeated argument that the process of language-learning grounds our language in experience. People use the word "blue" in their affirmations in situations where they have learned that people in general use the word "blue". This language-learning process is what supports the criterial relationship between S's affirmation "here, now, blue," and the protocol statement "S observed blue."

The relationship between a person's affirmation that they see blue and the protocol sentence that gets incorporated into science is a criterial one. The meaning of "blue" in the intersubjective language is determined by the conditions under which people as a group are disposed to use the word rather than the disposition of any individual, but the group is just the sum of its parts. We determine what things in the world are blue by when people in general are disposed to say "here, now,

blue," which means that affirmations contribute to the meaning of "blue" despite the meaning of "blue" not being *reducible* to the fact that people say they see it. This link between affirmations and the system of science falls short of entailment, but it is nevertheless stronger than merely psychological. Affirmations therefore provide not just motivation, but justification for acceptance of corresponding protocol sentences. That a person says "here, now, blue," is a criterion of it being true that they see blue, and in the absence of defeaters we should accept the protocol sentence into the system of science.

I think that this account is what Schlick was trying to get to when he published his account of the foundations of knowledge. His affirmations fit into Wittgenstein's criterial account, and his examples and claims about them are paralleled in Wittgenstein's writings and dictations from around the same time. As far as the history of Schlick's foundationalism goes, this is where I stop. Philosophically, however, I think that this account is flawed and incomplete, which perhaps contributed to Schlick's difficulty in getting people to understand or accept it. It is possible that Schlick was at this point just trying to put words around ideas of Wittgenstein's which were not fully formed at the time, and which never fully crystalized into a coherent account. I will elaborate on the objections I have to criteria in the next section, and in the next chapter we will explore whether it might be possible to achieve Schlick's goals in another way.

5.2.5 Problems for Criteria

Wright highlights the following initial objection to criteria. When we know that the criteria for *P* are satisfied, we are justified in asserting that we know *P*. But we have seen that criteria are defeasible, meaning that it is possible for subsequent evidence to contradict *P*. This subsequent evidence will not, however, have any bearing on the fact that the prior criteria for *P* were satisfied (it will instead move us to a point where *P* has different criteria). Therefore, knowledge that the criteria for *P* are satisfied is essentially different from knowledge that *P* – we can lose one without losing the other. So far, this is fine. The nature of criteria is that they *justify* knowledge claims without *constituting* knowledge claims – the criteria for *P* are not

the same thing as *P* (lying on the floor clutching a broken leg is not the same thing as being in pain). But, asks Wright, "[i]f recognition that criteria for *P* are satisfied can justify a knowledge-claim that *P* without *constituting* knowledge that *P*, what would constitute it?" (1984, p.385, original emphasis). For our purposes, we have proposed that Schlick may have been saying that affirmations are the criteria of protocol sentences, such that when an affirmation is made it will justify our acceptance of some related protocol sentence, but the protocol sentence can nevertheless be rejected later. If we later reject the protocol sentence, we don't thereby reject the affirmation. (I'm assuming we can make sense of saying that an affirmation that was made yesterday was certainly true at the time, even though we can't say what that affirmation was because we don't have access to the rules under which it was formed, given that affirmations have "no duration" (UFE, p.382).) But if knowledge of a protocol sentence isn't the same thing as knowledge of an affirmation, what is it?

Wright argues that either the truth conditions of *P* are accessible to us at the time we assert *P*, or they aren't. If they are accessible to us, there is no reason to use the notion of criteria instead of plain old truth conditions. If they are not, then there is scope for sceptical doubt to creep in. On this latter option, things will seem to be the same to the speaker whether *P* is true or not, and it is surely not the place of a grammatical convention like criteria to say that the things we do have access to entitle us to make claims about the things that we don't. This is not even to complain that there might be something inaccessible about, say, pain, that goes beyond the behaviour of people who are hurt. Wright points out that we could introduce a convention that says snowdrops by the end of January justify the claim that March will be dry (1984, p.386). This involves no notions of truths that are essentially beyond the reach of experience, just truths that are beyond the reach of our experience right now, so it is clear that we can make sense of the grammatically-determined criteria being satisfied without thereby getting a justified belief about that of which they are the criteria. We could limit our conventional criteria to only hold between things that we have already found to be linked, but relying on inductive reasoning like that is clearly just playing into the sceptic's hands.

Snowdrops and a dry March are independently identifiable, and therefore, given what we have said before, they would be better thought of as symptomatically related to one another (if at all) than criterially related. Criteria are appropriate for things that cannot be directly observed, such as a subject's ability to play chess or another person's being in pain, and this opens a possible response to this objection. Wittgenstein's point in discussing the criteria for knowing how to play chess is that there is nothing more to having this ability than making the right moves in the right places. Knowing that the right moves have been made is not the same as knowing that the subject has the ability to play chess, because no particular set of moves is identical with the ability, but this isn't to say that there is some further thing that we are failing to observe when we watch them make the moves. Wright's initial dichotomy is a false one – it is not the case that either the truth conditions of "S knows how to play chess" are accessible to us or they aren't, because there is no good notion of "truth conditions" for this proposition beyond the criteria. (This appears to be the line of response favoured by Hacker, who rejects formulating criteria as "necessarily good evidence" for precisely this reason (1984, p.314). He says there that the idea of there being a gap between evidence and that for which it is evidence is a confusion.)

Wright does consider the idea that criteria might replace truth conditions completely for at least some kinds of proposition (1984, pp.387-388), and he doesn't think it will help. He argues that if we can make predictions on the basis of the satisfaction of the criteria for something, as we surely can if the notion of *defeat* is in play (since if we didn't expect anything, nothing would run counter to our expectations), then we can certainly talk of *incorrectness* as associated with our expectations being disappointed. If I have observed S making permitted moves in a several consecutive games of chess, then the criteria for "S knows how to play chess" are satisfied, but we have said that this is consistent with later finding out that S was just getting lucky and subsequently having to reject the statement. In this case, wouldn't we say that we were wrong when we asserted that S knew how to play chess, even if we were justified in making the assertion? And now it once again looks like "S knows how to play chess" says more than just that up until now they have

made only permitted moves – it also says that they will make only permitted moves in the future. Maybe induction does confer justification, but that seems like the kind of substantive question that shouldn't be settled as a matter of grammar.

Bringing this back to our question about protocol sentences, what do we mean when we say, for instance, "S observed blue on Thursday," and how does it relate to S's affirmation "here, now, blue,"? If the affirmation is a criterion for acceptance of the protocol sentence, then we will be justified in accepting the latter once the former is asserted, but we might later reject it if we found out that, say, S had taken a large amount of LSD just before we arrived. This undercutting defeater changes the criteria for "S observed blue on Thursday" so that assertion of it is only justified if we have evidence apart from the affirmation. This is because we only want to say that "S observed blue on Thursday" can be true if there was a particular wavelength of light reflecting into S's eyes on Thursday. The truth conditions of the protocol sentence are in the public language and are linked to the conditions which have been found to obtain when people across the population use the word "blue", and someone who is likely to be hallucinating is a bad source of justification for claims about the physical characteristics of the world. What we know is that the subject has asserted that they saw blue. Why does that justify our assertion that a certain wavelength of light was being reflected at S's eyes, when that clearly goes beyond what "here, now, blue," really says? Criteria might be the only way of justifying claims about things to which we have no independent access, like pain or abilities – we don't have anything beyond the subject's actions to go on in these cases, and moving from *writhing* to *pain* might, as Hacker says, require no inference at all – but in the case of "blue" we *do* have independent access, and it is an empirical matter whether or not the subject's use of the word "blue" lines up with the right wavelength of light hitting their eyes.

Wright diagnoses the problem as follows: "The root of the problem on both horns is, obviously, the logical distinctness … of the states of affairs respectively constituting satisfaction of the criteria for, and the truth of, the relevant knowledge claim. … So the troublesome logical distinctness will have to be prevented. But how?"

(*ibid.*, p.386). *This is exactly the same problem as we have been grappling with for this entire chapter*. The problem with using criteria to try to bridge the gap between affirmations and the system of science is that Wittgenstein doesn't really say anything about how the relationship between *P* and the criteria for *P* is supposed to work. He says *that* there is a relationship, and I remain satisfied with the historical claim that Schlick was attempting to put Wittgenstein's ideas to work here, but philosophically this amounts to nothing more than an appeal to authority: Wittgenstein said that criteria could do what Schlick was looking for, therefore something can do what Schlick was looking for. Ultimately, however, it looks like discussion of criteria just collapses back into asking *how* these relata are supposed to be connected. Saying "it's a rule of grammar that they are," is just insufficient when we're looking for the grounds for significant epistemological claims. Because of this, I think that if we're looking for precise and certain foundations for science, we will need to look elsewhere.

5.2.6 A Private System of Science

We have argued that, if there is a logical connection between protocol sentences and affirmations, then it is impossible for affirmations to have a stronger epistemic status than protocol sentences. Since Schlick's position *was* that affirmations had a stronger epistemic status than protocol sentences, it was not possible for the two to be logically connected, and so we have argued that Schlick's model must have been based on a relation that was weaker than deductive. But we have also argued that this model is less than satisfactory. In this section I want to briefly consider one final area in logical space – could we adopt a view on which affirmations and protocol sentences are both infallible? This view was certainly not Schlick's, and I will argue that it is flawed in its own right, but it raises considerations that will be important going into the next chapter.

We know that the certainty of affirmations is supposed to be grounded in their demonstrative character, but also that their meaning can't just be "whatever it takes to make this proposition true" because the notion of truth conditions wouldn't apply in that case. However, in order to avoid the risk of using the wrong word to describe experience, Schlick said that the rules for interpreting an affirmation "lead

to ostensive definitions," (1935b, [1979b p.413]). This seems to imply that the rules for the interpretation of an affirmation "here, now, φ," include "φ is defined as 'that which is here and now," which looks like exactly the violation of private language concerns that we were worried about.

In chapters 2 and 3 (sections 2.3 and 3.1.2) we argued that, because of his Form and Content philosophy, ostensive definitions play a slightly different role for Schlick than might have been expected (perhaps to the extent that we shouldn't refer to them as "ostensive definitions"). We argued that, since meaning is logical form rather than content, ostensive definition provides a way of getting at the grammar of a word without relying on definition by other words, rather than attaching a word to a particular experienced sensation. In this way, ostensive definitions give us a kind of meaning which is publicly available – logical form can be grasped by two different people. However, in the case of affirmations the logical form of the ostensively-defined term is that of the experience of the definer. There is no guarantee that anyone else present has the same form of experience. This means that affirmations fall afoul of our first private language concern from pages 141-142 – the language thus defined does not necessarily line up with the language everyone else is speaking. This appears to make affirmations useless for communication, but we might be able to use them to confirm an individual's system of beliefs if we can avoid the second objection – that a statement the meaning of which is *whatever makes this statement true* isn't a meaningful statement. For this to be the case, the speaker must attach a determinate meaning to the words of the affirmation. The meaning of the words can be chosen to correspond with what the speaker believes to be the form of their experience, but the speaker also has to understand that if the world was different then the same affirmation would be false. If that condition is met, then the affirmation is not logically infallible, but we may consider it to be epistemically infallible if we are prepared to accept, as Schlick did, that we can select rules of interpretation for our utterances that we know to correspond with what we are experiencing.

Affirmations thus construed can be linked to the intersubjective language because, within the Form and Content framework, there is no possibility of a language which isn't intersubjective in this sense. Language is about form which can be shared by sentences, minds, and the facts which they express, and there is no distinction between the material world that is the subject matter of science and the internal world of experience. A person has a number of hypothetical beliefs about the world, such as "when I look through the telescope, I will see red," and they form affirmations in the same language. When the subject looks through the telescope and says "here, now, red," the way that this affirmation confirms the hypothesis is that "red" means the same thing in the affirmation as it does in the hypothesis. At the same time, however, "red" in the affirmation is defined ostensively with reference to whatever the subject is looking at. The implication of this is that what the subject believes about the world – the hypothetical belief here being confirmed – is determined by what they are looking at. When a person looks at, say, Mars and says "red" they do so because they believe that the word "red" is the appropriate one to use to describe this colour. They believe that the colour they are now seeing is similar to that of post boxes and contrasts with that of grass. The result is that, on this account, the meaning of not only the affirmation, but the whole network of the subject's beliefs is determined by the ostensive definitions which occur moment by moment in the forming of affirmations.

Affirmations are statements about the world, and when they are uttered sincerely the subject will be choosing the words which they believe are right to describe their present experience. Meaning, on the Schlickian account, is a matter of the rules chosen for the interpretation of signs, and so when a subject sets out to describe whatever is in front of them, whatever that is (but not *whatever that is*, which we have seen will fall afoul of private language concerns) becomes the meaning of the term used. But the subject isn't just using some placeholder term like "this" or "φ" – they're using a word that has a place in the rest of their beliefs (and even if they can't find the right word for what they're currently experiencing, they're still trying to refer to something which bears a relation to the rest of those beliefs – even if I can't think of the right word for a colour, I can point to it on a colour wheel).

A person who affirms "here, now, blue," at the place where they were expecting to see blue is confirming their beliefs, and that just goes to show that *what they believe* is *this*. If affirmations are to serve as foundations of knowledge with both infallibility and a deductive link to the system of science, then it looks as though the result must be that a person's entire network of belief is given meaning by the moments of experience which verify or falsify them.

This view might hold some appeal to a radical internalist who is happy enough for the entirety of their language to have its meaning determined moment by moment by their current beliefs. The tension here is that, in believing that "blue" is the *appropriate* word to describe their experience, the subject seems to bring more to the meaning of the word than we would expect from a simple ostensive definition. They bring along with them beliefs like "this word is also the word which applies to such-and-such", and "this word is the word that most people would use to describe similar situations". Certainly we think that there is some stability in the majority of our hypotheses about the world, and that would seem to be enough to reduce this account to absurdity, but the problem is present even in the view that we attributed to Schlick in the previous sections. Schlick wants it to be the case that the sense of an affirmation is not determined by common linguistic practices, and that is what we need if we are to retain certainty, but in this case why try to use common words in expressing them? Why don't we make up a new word for every new moment of experience? It can only be because we have certain beliefs about the way these experiences link in with our pre-existing beliefs, and that introduces epistemic risk.

So, if the system of science as a whole is, as we assume, fallible, then if affirmations are infallible our use of words like "red" and "blue" is inconsistent across the language. What we have just seen in this section is that if we hold our use of those words as fixed across the system, and still take affirmations as infallible, then this seems to result in a subject's language shifting moment by moment in accordance with whatever they're currently experiencing. That means that in either case we have a troubling lack of consistency in our use of words whenever

affirmations are held as infallible in the moment of utterance. In the next chapter we will see how the problem of secure foundations for science evolves when we get clearer on the stability of meaning. We will see how Donald Davidson provided an account of language on which we must take it as primitive that people are using language in some systematic way, and we will go on to look at the framework provided by David Chalmers within which we can interpret meaning-change over time with a bit more precision.

5.3 Conclusion

We have seen that Oberdan was right to emphasise rules of interpretation as crucial to an understanding of Schlick's affirmations, because an affirmation necessarily involves a demonstrative term. However, Oberdan further understood Schlick to have been talking about experience in an intersubjective language in the final part of the affirmation, which makes it impossible to guarantee the statement's truth. Instead, Schlick must have conceived of affirmations as being formed in a language the meaning of which is entirely transparent to the speaker. As Uebel identified, in an affirmation Schlick allows the speaker to select their own rules of interpretation at the moment of utterance, and this prevents us from saying that the affirmation bears directly on the claims of an intersubjective science. It seems as though Schlick was attempting to make use of Wittgenstein's emerging notion of criteria to give affirmations *some* link back to the system. A person's report of first-person experience would satisfy the criteria of a corresponding protocol sentence and thereby justify its inclusion in the system of science, but this justification is defeasible and the protocol sentence itself can be rejected later for any number of reasons, though this rejection would not mean that the affirmation was *wrong*, per se, just that it had been incorporated incorrectly. However, Wittgenstein's account of criteria does little to explain why we would be justified in accepting a claim on the basis of something that is not inferentially connected to that claim, and for that

reason I have to conclude that Schlick failed to give a complete explanation of the link between affirmations and the system of science.

The goal now is to see whether there is anything that might fulfil the objectives Schlick set out to pursue: for affirmations to be epistemically infallible *and* have a role to play in a grounded empiricist system. I have suggested that rather than being merely psychological motivators or mysterious justifiers, there is room for an interpretation of affirmations as saying something about a real feature of the world. They bear appropriate relations to the intersubjective system of science because, as we have seen from Schlick's Form and Content arguments, all statements are intersubjective, and from them we verify or falsify the rest of our belief network. The cost of this approach is radical internalism – the requirement that the system of science be based on these affirmations leads us to a state in which the meanings of beliefs within the system of science itself are determined by the same rules of interpretation as the affirmations, selected moment-by-moment by the speaker. The problem which now presents itself is that the language selected has no principles of consistency from one moment to the next. It is clear that there is some reason that we choose to use the word "blue" to describe certain experiences. We have beliefs about which words are generally the correct word to describe a situation, and if these beliefs have a bearing on the system of science then they will also creep into the meanings of affirmations. These beliefs seem to go beyond immediate experience and, if this is the case, thus throw the affirmation's infallibility once again into question. In the next chapter we will explore two accounts of language which were formed in the years since the protocol sentence debate. Davidson's account will bring into stark relief the worry that the consistency of language-use is an issue for a private language account of affirmations and science. However, Chalmers' scrutability framework may give us the tools we need to evade this worry and find a way to ground the system of science in experience.

Chapter 6 - Coherence and Scrutability

6.0 Introduction

In this chapter, I will talk about the way the questions raised in the previous chapters have been addressed since the protocol sentences debate. The goal is to see whether the more recent contributions help to show us where, if anywhere, Schlick went wrong, and how much of his project might be rehabilitated into a coherent position. The position laid out will not be Schlick's, but will aim to ground scientific knowledge in experience with epistemically secure foundations.

Firstly, we will look at Donald Davidson's account of truth and knowledge, which he develops drawing on lessons from Schlick and Neurath's debate. Davidson attempts to give an account of the interrelationship of belief and truth, such that we have reasonable grounds to think most of our beliefs true (although no single belief will have firm epistemic standing), despite the fact that we only compare our beliefs with other beliefs and never with reality directly. Davidson's work might be seen as an account of how we can establish the overall reliability of science without establishing firm foundations at any point. I will argue that Davidson makes important points, such as focussing on what we can learn about language as interpreters rather than as speakers, but Schlick's arguments for the importance of first-person reports of present experience are not entirely defeated.

Secondly, we will look at the recent work by David Chalmers on philosophical construction, which takes a fresh look at Carnap's *Aufbau* and considers various ways in which we might adapt it to deal with subsequent criticisms, whilst still retaining the spirit of the original. We will see that there a significant differences between Chalmers' own account and the view that we have attributed to Schlick, but the framework within which Chalmers develops his view is instructive, and at the major choice-points which he identifies for alternative internally-consistent views it is clear enough which choice a Schlick-inspired account will take. I will argue that

the system of science as a whole appears secure, but this cannot be the result of security in a base of individually-infallible experience statements.

6.1 Coherence and Evidence

In this section we will explore Davidson's coherence-based account and see what we can learn from it. First, I will ask what we mean by "evidence", showing how the debate between Schlick and Neurath thus far has been about what kinds of entity can fill this role. Then I will talk about Davidson's argument for the reliability of coherent systems of belief which leads us onto epistemic externalism – a position which I will consider in further detail in section 6.1.3. In section 6.2, we will return to Schlick and ask where the differences and similarities lie between Schlick and Davidson's positions.

6.1.1 Evidence

The dispute which arose between Neurath and Schlick was, at base, one about the nature of evidence, and Davidson raises the question anew in "Empirical Content" (1982, [2001]). Is evidence an object, a fact, an event, an experience, a sensation, a sentence, a proposition, or a belief? For example, suppose we are considering the general proposition, "all ravens are black,". We might in this case regard a particular black raven, an object, as a piece of evidence for the claim. But it is more plausible to think that the evidence is the *fact* that the raven is black, because the same raven could conceivably exist in a non-black state (if ravens go grey in old age, or if it was dyed by someone), in which case the same object would no longer count as evidence for the claim in question. In this non-black-raven case, the object still exists but the fact does not. Given that the evidence does not exist in this case, the object obviously has different properties to the evidence and it is therefore not identical with the evidence. So, evidence is not an object, but it may (on the basis solely of these considerations) be a fact.

The difference of opinion between Schlick and Neurath is over where we go from here. Schlick thought that there were at least some propositions for which our

evidence was a matter of fact. Neurath thought that the only possible evidence for a proposition was another proposition. And we have seen already (and Davidson reiterates (1981, [2001, p.146])) that in the latter case we are left with a coherence theory (not yet distinguishing between a theory *of* truth or knowledge) and all of the associated problems, whereas with the former case it at first appears impossible to defeat a generalised scepticism.

Davidson argues that there are important insights to be gained from each alternative. It must be acknowledged that not just *any* set of coherent beliefs can be called accurate. If "true" means anything at all then it surely denotes some kind of correspondence with an external reality. A coherence theory of truth allows an infinite number of mutually incompatible sets of truths, which makes nonsense of the word "truth". At the same time, it seems as though the only evidence we can ever appeal to in the justification of our beliefs is another belief. Thus, Davidson argues, if we want to claim that we have knowledge about the external world we should look for a reason to think that a coherent network of belief should, under normal conditions, correspond with the way the world is (1981, [2001, pp.137-138]).

6.1.2 The Reliability of Coherence

Davidson adopts a Quinean account of language, on which there is no analytic/synthetic distinction and understanding a language is holistic rather than a matter of understanding individual statements. *Assent to a sentence* plays a foundational role in the determination of meaning here, rather than the other way around. This allows us to fix assent to a particular sentence in the face of apparently-contradictory evidence. If a subject is prepared to assert "all ravens are black," and is then presented with a white raven, then they don't necessarily have to reject the original claim – they can make changes elsewhere in their network of beliefs and assert, for example, "I'm currently hallucinating". From assent to sentences, which is publicly available information, we as interpreters have to try to work out what the speaker believes – this is essentially the point of communication. What a speaker believes obviously is connected to what they assert by what they mean when they assert things; as Davidson points out, "a speaker's assent to a sentence depends both on what he means by the sentence and on what he believes about the world," (1981,

[2001, p.147]). The problem now is that there are two unknowns – meaning and belief – and only one fixed point from which to try to derive them.

Say we are trying to interpret the behaviour of a speaker so as to attribute beliefs to them. We can generally know whether or not a speaker assents to a sentence by their behaviour, and we assume that someone who wishes to play a part in the same language game as us does not systematically deceive us, but we have no immediate access to either what they take the sentence to mean or to what they believe to be the case about the world – those are the things which we are trying to find out. Davidson's solution to this problem is that we must adopt a principle of charity and apply our own standards of truth to the speaker. We have our own beliefs about how the world is, and we need to assume that the speaker shares at least some of the most straightforward beliefs about the immediate environment. We must hold one of our variables constant so that we can figure out the other, and the thing which we must hold constant is belief. We must assume that the person occupies the same world as us and believes the most straightforward things about it. Then we will know that whatever he says should line up with the way the world is in the right way.

The implication of this is that a correctly-informed interpreter will always come to believe that a speaker is (at least mostly) correct about their beliefs. A speaker who was thoroughly mistaken about how things stand in the world would not be able to express those beliefs, because frequently the sentences to which they assented would mean something to them which could not be found by an interpreter in the immediate environment. Thus, a charitable interpreter would have to interpret the sentences as having a different meaning than that which was intended by the speaker. So, it will always transpire that the majority of the beliefs which an

interpreter attributes to a speaker are true and mutually consistent (1981, [2001, pp.150-151]).[31]

6.1.3 Epistemic Externalism

Davidson next asserts:

> What stands in the way of global skepticism of the senses is, in my view, the fact that we must, in the plainest and methodologically most basic cases, take the objects of a belief to be the causes of that belief. And what we, as interpreters, must take them to be is what they in fact are. Communication begins where causes converge: your utterance means what mine does if belief in its truth is systematically caused by the same events and objects.
>
> - Davidson, 1981, [2001, p.151]

Davidson's claim here is that whatever object in the world actually, systematically, causes our beliefs is the object to which our beliefs refer. So, for example, in order to successfully be able to refer to tables it doesn't matter what the content of your phenomenal experience is when you're looking at a table or whether the thing in the real world actually corresponds in any way to what you think a table looks like, so long as, by some causal mechanism, it is tables which systematically prompt you to assert (or assent to assertions of) "there is a table in the room". Davidson argues for this claim in "Epistemology Externalised" (1991, [2001, p.195]) by pointing out that if the object of a belief was determined by our other beliefs about what we think is the subject (i.e. if the reference of "table" is determined by a

[31] Davidson notes at this point that his view differs from Quine's by associating meanings with facts in the world which prompt assent, rather than with sensory stimulations. Thus, Quine is committed to a distinction between factual statements and observation statements, which Davidson is not. Quine's distinction has epistemic significance in his philosophy – observation statements are evidence, in some capacity. Davidson has argued above that the only evidence for a belief is other beliefs, which is why he makes meaning into something about the external world instead.

set of descriptions which we believe to be individually necessary and jointly sufficient for an object to be a table) then our concepts would change every time we formed a new belief. Even if we distinguish between analytic and synthetic truths (which Davidson does not), it is clear that empirical data informs at least *some* of our definitions.[32]

Davidson has another argument which tacks along the familiar line of language-learning, which we saw in chapter 3 was behind Schlick's argument for the verification principle (see section 3.1.2). Schlick argued that the only way to teach someone a language who did not already know one was through ostensive definitions, which is to demonstrate to them the conditions under which the sentence was verified. Schlick concludes that the only meaning it is possible to learn for a statement is that which you can get from the sense data associated with that statement's truth. Davidson argues that communication takes place in a three-part system. There is the world and two speakers who are talking about it, each of whom can see and respond both to one another and to the world. In the case of language learning, one speaker takes on the role of language teacher and the other that of language learner.

> The learner is rewarded, whether deliberately or not, when the learner makes sounds or otherwise responds in ways the teacher finds appropriate in situations the teacher classes together. The learner is subsequently caused to make similar sounds by situations the learner instinctively classes together.
> - Davidson, 1991, [2001, p.203]

[32] This fact was well-known to Schlick from the time of AE (pp.46-47), and we addressed it in chapter 1 (see pages 17-18): different judgements in which a concept features serve as definitional of that concept at different times. It may also be worth noting at this point that various philosophical projects being carried out today operate under the assumption that concepts can shift significantly over time without losing their identity (e.g. Haslanger, 2000; Burgess & Plunkett, 2013; Cappelen, 2018 (forthcoming)), possibly shifting both intension and extension.

By a repeated process of correction and reward, language learners and teachers are caused to respond to the same classes of situations in the same ways, and thus the contents of the responses are fixed by the classes of things which cause those responses.

The notion of "cause" may at first seem to introduce some difficulty to Davidson's position. If I come to believe "the sky is blue" in the most basic way possible (i.e. by looking at it and having certain prior beliefs about what "blue" means and where the sky is to be found) then it seems as though the sky is causally connected to my belief. But it has only a tenuous claim to being *the* object that is the cause of my belief. Electrical impulses through my optic nerves are objects in the world, and they play a crucial role in the process. Nevertheless, the object of my belief "the sky is blue" is not my optic nerve. We might want to say that "blue" says something about the impact of light on my eyes, but that won't do because "the sky is blue" remains true even if I go blind. The sentence is (apparently) about a publicly-accessible object, not individual sensations, but it's the sensations which cause my belief. Davidson's reply to this is that whatever the cause is, it must be something which is likely to cause utterances from others. If other speakers and I are to refer to the same thing by "sky" then I shouldn't stop tracing back causes when I reach my optic nerve. "Communication begins where causes converge," (Davidson, 1981, [2001, p.151]) so the object which is referred to by "sky" is that object which is causally linked to both of our utterances of "sky" on the various occasions when we're inclined to so utter it – that is, the sky.

The role of the externalism in Davidson's account is to make it impossible for *all* of one's beliefs to be false (although any individual one may be called into question). "If anything is systematically causing certain experiences (or verbal responses), that is what the thoughts and utterances are about. This rules out systematic error," (Davidson, 1991, [2001, p.201]) meaning that our network of beliefs must be mostly-correct because the meanings of the terms in which the network is expressed are fixed by the way the world actually is in whatever manner makes it so that our beliefs are mostly accurate.

6.2 A Comparison of Schlick and Davidson

6.2.1 The Correspondence Theory of Truth

As noted, Davidson argues that the coherence theory *of truth* is deeply flawed, so if the alternative is a correspondence theory then we can commit to it. We have seen that Schlick's emphasis on the correspondence theory of truth is a common theme in his work from his very earliest papers right through to UFE (see chapters 1 through 5). We have also seen that the arguments which he raised (though he was by no means the first) are accepted even by those who wish to espouse a coherence theory of belief.

The central claim of the correspondence theory of truth can be expressed disquotationally, as famously occurs in Tarski's work: "P" is true iff P. The interesting work from Tarski is in showing that we can make sense of a distinction between object language and metalanguage and can use one to talk about the other – an idea which also shows up in Schlick's "On Facts and Propositions" with his claim "words denoting symbols and words denoting other things may occur in the same sentence" (1935a, [1979b, p.402]), as was pointed out by Thomas Oberdan (1996, p.285). Schlick's claim is, and always has been, that a statement is true (or a judgement is accurate) if and only if there is a fact in the world which corresponds to it.

We saw that Schlick's position here underwent some evolution between his 1910 work and AE (see footnote 9 on page 21). Originally Schlick said that there had to be a one-one correspondence between facts and statements, so that each fact in the world made exactly one statement true and each statement was made true or false by exactly one fact. We saw that Schlick continued to maintain that each statement must correspond to exactly one fact (which might be a disjunctive fact – "P v Q" is true iff P v Q), and that seems to be entailed by a disquotational definition of truth (that's the role of the "only" in "if and only if"), but later Schlick allowed for the possibility that the same fact would make more than one statement true. This is certainly correct, since the single fact that is the entire state of the world will imply

every true statement, and it is not the case that all of these statements are equivalent to one another.

If the only competitor to the correspondence theory of truth is the coherence theory of truth, then the correspondence theory of truth is the definition of "true" we must adopt, because the coherence theory of truth leads us to the nonsensical conclusion that we can choose between infinitely many mutually incompatible systems, all of which are equally true. Davidson, in "A Coherence Theory of Truth and Knowledge", does not defend the coherence theory as an account of truth, but rather defends an account on which coherence of belief *entails* truth by correspondence. And we have argued (section 4.4.4; see also Uebel, 1992, p.270) that even Neurath was not attempting to defend a coherence theory of *truth*. Neurath was attempting to give a coherence account of belief, or acceptance into a system, in a similar way to Davidson, as an empirical claim about how science actually progresses. Neurath's view of truth was that it was a metaphysical concept which we should replace or purge from our language. (In fact, Tarski can also be interpreted as claiming that truth is essentially unanalysable (Putnam, 1960, [1975]), which is to say that disquotation doesn't really do anything to explicate the concept in the way that, say, a picture theory of the correspondence relation does.)

What we may conclude is that, *if* we can make any sense of a notion of truth, that notion is a property which statements have whenever they correspond to how the world actually is. What Neurath got wrong was thinking that we could do without truth by moving to a holistic account of belief. Because even if it's right that we can never finally verify the truth of even simple observation statements, and even logical truths might be rejected on review, we still need an account of what it *is* to believe or reject a statement. What does belief aim at if not truth? When a person rejects an observation statement or a logical claim, what they come to believe is that "such-and-such" is false. So, we cannot make do without truth to serve as a target for belief, and the only candidate for a definition of truth is correspondence, so we must conclude that a statement is true iff the fact that corresponds to the statement obtains.

Schlick's correspondence theory of truth is different to Davidson's in the way the correspondence is set up. Schlick endorses Wittgenstein's picture theory of language, as we saw in chapter 2, which means that the elements of the sentence are arranged in a particular way and the fact which corresponds to the sentence is the fact that the objects referred to by those elements are arranged in the same way. The correspondence relation set up by Davidson, on the other hand, is, as noted, causal. The sentence corresponds to the fact that systematically prompts assent to that sentence. In this case, the correspondence theory alone does not impose the condition that there are as many elements of the sentence as there are elements of the fact. Davidson in fact argues that it is a confusion to ascribe the same number of elements to the fact as to the sentence. In the case of the expression "the father of Annette," for example, on the picture-theoretic account we saw that we needed to posit an entity which was the relation *xRy* such that "*aRb*" is true iff *a* is the father of *b* (see section 2.1.2). But Davidson argues that for any expression which consists of prefacing "Annette" with 0 or more repetitions of "the father of" it will be straightforward to give a theory which points to a single entity to which the expression refers:

> [I]f the term is 'Annette' it refers to Annette, while if the term is complex, consisting of 'the father of' prefixed to a singular term t, then it refers to the father of the person to whom t refers. It is obvious that no entity corresponding to 'the father of' is, or needs to be, mentioned in stating this theory.
> - Davidson, 1967, p.305

Instead of having each part of a sentence refer to an object, Davidson believes that it suffices to have a rule that explains how that sentence-part systematically transforms the meaning of the sentence in which it occurs. In part we can put this down to a merely verbal dispute about what constitutes an 'entity', since there are obviously relations between things and we can obviously talk about them, but it does give Davidson's account more flexibility to deal with problems such as the one we encountered in section 2.1.2 regarding the difference between the relations *aRb* and *bRa*.

6.2.2 Speaker Meaning and Interpreter Meaning

An important point is raised by Davidson that is not considered significant by any of the earlier writers we have looked at, which is the difference between meanings as assigned to a sentence by the speaker of that sentence and meanings as assigned by a well-placed interpreter trying to understand the sentence. Given that language is a tool of communication, it seems reasonable that the understanding of interpreters should be the focus of a theory of meaning. It doesn't matter if two people have radically different things going on in their heads as long as they can, e.g., give each other directions and not get lost.

I do think, however, that Davidson errs in the implementation of this idea. We said, in section 6.1.3, that the object that is taken to be the cause of our beliefs should be the public object that naturally fits into how we divide up the world. We can't say that the meaning of a term is a particular kind of firing of my optic nerves, for example, because certainly no one *else's* use of the word refers to *my* optic nerves. But, it's equally clear that my optic nerves, and subsequent state of mind, can be manipulated in the same way by more than one cause. In everyday cases, the glass of water on my desk looks the same when the water is replaced (which is perhaps why I don't name my individual glasses of water), and in more elaborate scenarios I might be placed in a particularly good simulation. Internally there is no difference between someone who has lived their whole lives in a simulation and someone who has not. Consequently, if we have someone who has lived their life and learned their language inside a simulation, and then we take them out of the simulation, they will now assent to the same sentences as us in the same situations even though their beliefs are now being caused by real objects, whereas previously they were caused by lines of computer code. If we assume that the simulation was a computerised version of the real world (perhaps minus an infinite nesting of people living in their own simulations), then this person could give us directions to a train station that they've never truly been to and we could follow those directions without difficulty.

When Davidson considers cases like this, he claims that the person leaving the simulation won't *really* be saying anything. His own example is of Swampman – an identical replica of Davidson created when lightning struck a swamp, around the

same time as the original Davidson is struck by lightning and disintegrated. Swampman will walk out of the swamp and behave exactly as Davidson himself would behave, because he has the same structure all the way down to the brain and mental states, but when he sees Davidson's friends and waves to them, Davidson claims, he does not truly recognise them because he doesn't have the correct causal history – "it can't *re*cognize anything, since it never cognized anything in the first place," (Davidson, 1987, [2001, p.19], original emphasis). Davidson takes the Swampman case to show that "people who are in all relevant respects similar … can differ in what they mean," (1987, [2001, p.32]), and it might be that some people share this intuition. However, it highlights a deficiency in the Davidsonian account: If Swampman doesn't share a language with the people he speaks to, then what *does* he share with them that allows him to communicate his ideas and allows them to learn truths from him? When Swampman tells us things, we learn. When he lies, he makes a meaningful utterance which is false, and we judge him for it. Perhaps we need to introduce a concept of quasi-language, which is the disposition of a subject to use particular words and gestures to communicate its ideas, regardless of their causal history. Swampman, or a subject who spends the first half of his life in a simulation, speaks the same quasi-language as us because we can communicate ideas to one another. Perhaps Davidsonian language is an interesting and unique sub-concept of quasi-language that can be studied in its own right, but it's clear that there are philosophically-interesting things to investigate about quasi-language broadly-construed. That's what I'm interested in, and to save space I'll just be calling it "language" everywhere outside this paragraph.

Given that a subject can learn a language inside a simulation and then be a competent language-user outside it, if we want to integrate Davidson's insights into a picture of language, we should look for something else to play the causal role. We still obviously don't want the meaning of an expression to be the firing of our own nerves, but it clearly needs to take into account the role that mental states play in the determination of meaning. There is room here for a certain amount of synthesis with the account we have been developing of Schlick's position. Let the cause of a belief be the structure of our experience. Davidson's insight is that as interpreters

we must assume that neither we nor our interlocutors are systematically deceived about the state of the world. Schlick's insight is that for this to be the case is for us to assume that the structure of our experience maps onto the structure of the world and that the same is true of the interlocutor. In the most methodologically basic cases, the cause of our beliefs is a particular structure to experience, which is something that can be shared by both speaker and interpreter. The modification being made to Schlick's position at this stage is that we don't get to impose our own rules on the way our language expresses that structure. The meaning of our assertions is going to be linked to the structure which typically causes those assertions – a structure which is shared by various different things-in-themselves that might cause it (reality, simulation, evil demon, etc.), but which is not arbitrarily chosen by the speaker at the moment of utterance.

6.2.3 Affirmations

Suppose that we broadly adopt Davidson's picture of belief and knowledge but with the change that the causes of our most basic beliefs are identified with structures of sense data rather than external-world objects.[33] Much of the picture we built up of Schlick's philosophy remains unchanged. We have a network of beliefs, or accepted statements, that needs to be internally consistent. As a hypothesis of empiricism, all meaningful beliefs are either analytic or mean what they do in virtue of their link to sense data. We have a number of beliefs about what sense data were given at various times and places and all of that is fallible as usual. The point at which we'll be looking to preserve infallibility, to preserve the spirit of Schlick's enterprise, is for the belief about what is currently being experienced. Of course, in a holistic system any one belief can be held constant with enough changes around it, but (i) we are also committed to preserving analytic truths, and (ii) ideally, whatever guarantees the infallibility of the present-tense first-person belief will not work equally well for holding some arbitrary protocol sentence constant,

[33] This is perhaps closest to the account of Quine.

because if it did it would make affirmations no more interesting than those other beliefs about experience.

For a general experience sentence, "E was experienced at point P at time T", our hybrid version of Davidson's externalist account says that "E" refers to whatever experience would normally cause us to believe that E had been observed, "P" refers to whatever point P picks out in our usual spatial reference frame (and is the position of the observer rather than whatever object is being perceived, since E is sense data and occurs either in the mind or in the sensory organs rather than out in the world), and "T" refers to whenever T occurs in our usual temporal reference frame. An affirmation is a belief of the form "E is here now," where "here" and "now" are to be linked to the system of belief by the other beliefs "I am currently at P," and "the time is T," which certainly carry no guarantee of accuracy. But neither can "E is here now," if "E" refers to whatever sense data normally cause us to say "E is occurring" because it seems that we don't have any guarantee that the same sense data as usual are what's causing us to think "E".

If "E" in a protocol statement refers to whatever normally causes me to assert E, it seems like "E" in an affirmation should refer to whatever causes me to assert E right now. In that case it's just another way of saying "this". At that point affirmations no longer have a rational relation to the system of science, because we can't make a logical deduction from "that which usually causes me to say 'E'," to "that which is now causing me to say 'E'." It looks like this is the point at which the Davidsonian account makes the most significant divergence from Schlick's, because his response to scepticism is to show that it couldn't be that *all* of our beliefs are false rather than to show that some specific class of beliefs is guaranteed. In the next section I will further elaborate the problem which is caused for affirmations by the requirement that the meaning of a term is that which is usually the cause of it being uttered.

6.3 Consistency

Suppose we're attempting to interpret the utterance of someone who says "rouge"

every time an element from a certain range of colours is presented to them, and never when in the presence of anything other than a sample of that colour-range. It seems right to say that in one way or another it's the colour which we are presenting to them which causes them to utter "rouge", and we seem able to conclude that "rouge", when uttered by the subject, refers to things of that colour. Next, the subject is shown something inside a box which we, the interpreters, cannot see into from our angle, and says "rouge". It seems right for us to conclude that whatever is in the box is an element of the colour-range which the subject has previously picked out. What is it that justifies our inference from the two known facts, "the subject has previously uttered 'rouge when and only when presented with elements in such-and-such colour range," and "the subject is now saying 'rouge'," to our presumably-reasonable conclusion "the subject is now presented with elements in such-and-such colour range"?

This is a form of the problem of induction, familiar from at least as early as Hume (1777, [1975, pp.32-35]). We know that strictly speaking the inference is invalid without the inclusion of some kind of implicit inductive premise, something like "subjects who have always and only uttered 'S' in the presence of φ in the past will always and only utter 'S' in the presence of φ in the future,". What justifies the inductive premise? In this case it seems to be that it is embedded in the assumptions which we must make as interpreters. We must assume that the person we're speaking to wants to be understood, and therefore does not wilfully sabotage the conversation. One way of them doing that would be, as noted, for them to lie, and we as interpreters must, necessarily, assume that they don't do so systematically. (It's not even clear that we can make sense of the notion of someone deceiving us systematically. We would just interpret them as speaking a different language.) Another way would be for the meaning of their words to change after every utterance. We might still be able to understand the subject, eventually, if the words changed according to some pattern (e.g. for the first instance of one of "A", "B", or "C", "A" refers to α, "B" refers to β, "C" refers to γ, for the second instance "A" refers to β, "B" to γ, and "C" to α, and so on, so that "A & A & A" contains no redundancy but "C & B & A" does), but suppose the subject chose words completely at random to

express their thoughts. It would be impossible to interpret the speaker because we would never be able to identify a pattern of rules which explained the subject's utterances. Even a single utterance would be unintelligible because we wouldn't know if the subject was picking out the shape of the thing in front of them, or its colour or smell or whatever, and there would be no hope of creating a translation manual. In this case, we would surely conclude that the speaker didn't really have a language at all. They just speak gibberish.

What the above example shows is that one of the requirements of language is consistency. There must be some rules that allow us to infer the meaning of some sentences from the known meaning of others (even if these rules must be loosely formulated, include provisions for vagueness, perhaps change over time, and can never be known with any degree of certainty). The idea that speakers and interpreters can each observe the world as well as each other's behaviour is useful for understanding how language can be learned in the first place, but it is clearly not the case for every instance of language use. Indeed, any utterance made with the objective of informing the listener about something they don't already know is going to involve a listener who can't immediately compare the speaker's utterance to the way the world is. If the speaker wishes to be understood, they must believe that they are making their utterance in accordance with the same set of rules that informs the rest of the utterances in their linguistic community.

On the account which we have been considering, the meaning of a term is determined by the conditions under which we are generally inclined to use the term. The rules for utterance of "rouge" are therefore something like *do so whenever referring to something that is red*. This rule might not be known, in this form, to the speaker, but could rather have been derived by an interpreter who could see red things and hear the utterance of "rouge". When the subject utters "rouge" again in the future, they must do so in the belief that they are following the consistent rules of their language, whatever those might be. They intend to use "rouge" the same way they used it last time. It's the same word. It has the same meaning. We can therefore infer that when the subject uses those words again, they use them in the belief that

the same relevant conditions obtain. When someone uses "rouge" it is because they believe that the proper conditions are met for them to utter "rouge".

Schlick's specific claim in "On Affirmations" is that there is a certain kind of sentence that is rendered nonsensical by the addition of words like "possibly" or "perhaps", and these sentences are something like "here two yellow lines coincide" and will be called "affirmations". Supposing we observe a subject who utters the phrase "here two yellow lines coincide", is there an interpretation of that sentence upon which it communicates something to us and yet in some sense or another cannot be mistaken, or at least cannot be *doubted*? In light of our considerations from a moment ago, the subject must believe that the present conditions are those under which it would normally be correct to utter "here two yellow lines coincide". But we could always add a "possibly" to something like that, as noted. Even if we think that the conditions are determined by sense data rather than the presence of something with particular physical characteristics, and thereby allow hallucinations or illusions to suffice as proper conditions, there is no guarantee that the same kind of sense data is the cause of the subject's utterance as normally causes such an utterance. There is no valid inference from "this is causing me to say 'E'," to "this usually causes me to say 'E'." Nor does it look like we can simply add "I believe that," or "it seems as though," because we can imagine situations in which a person has overwhelming evidence that their senses are not to be trusted and so it doesn't even seem to them as though what they are experiencing says something about the world. On the other hand, we saw in the last chapter that experiences are just as much matters of fact as the world they represent, because facts are expressed by facts. Affirmations are clearly supposed to be sentences about experiences and not sentences about the world.

Here is Schlick's example from "On Affirmations" in fuller detail:

> Suppose a physicist wants me to check some project. He gets me to look through a telescope in his laboratory and asks: 'What is there now in the visual field?' I answer (truthfully, let us assume): 'There are two yellow lines there.'

…

> In our case it would be absurd to say: 'Perhaps there is yellow in the visual field'. If I gave such an answer to my physicist, he would tell me: 'You must have misunderstood me. My question, as it was intended, cannot be answered with a "perhaps".'
> - Schlick, 1935b, [1979b, p.409]

Schlick goes on to consider three ways in which a "perhaps" might enter into the answer, but which turn the sentence away from being an affirmation. The first is a case in which the speaker is talking about the world rather than their experience of it. The other two possible sources of doubt both concern the use of the word "yellow", and we will see that the latter of these is genuinely problematic.

Source-of-doubt number one is the case in which the speaker has reason to think that there is not *really* yellow in the world – that other observers of the same thing would not say "yellow" because of illusions or hallucinations unique to the speaker's situation. If this is the reason for the subject's doubt, then it is not an affirmation because the speaker is talking about the wrong kind of thing. We are interested in the facts of their experience and not the facts of the world. Thus, Schlick says, this does not cause a problem. We might worry about how we can direct our reference to experience rather than to that which experience is about, given that we have looked at how language is learned and seen that it always leads reference to things that can be observed both by learner and teacher, and so reference to the stuff of experience seems impossible. However, for the sake of argument, let's suppose we can set up such a reference with phrases like "I'm now having an experience as if the world is in such-and-such a state." We will see that this won't get around the problem with source-of-doubt number three.

The second source of doubt centres on the vagueness of language. If we're looking at an area where yellow merges into green, for example, then it might need to be clarified that the word "yellow" has a vaguer meaning than usual, but Schlick thinks this does not affect the accuracy of the affirmation itself, and this seems reasonable. Vague statements can be true.

In the last case, the source of doubt was that the speaker could not be sure they were using the word "yellow" in the way they generally do within their linguistic community. This is exactly the issue we raised at the beginning of this section (and the end of the last chapter) – that when we speak, we do so on the assumption that we are following the consistent rules of our language, and if we have reason to doubt that then we should be less confident in our assertions. Schlick's response seems to draw a distinction between uncertainty in our sentence and uncertainty in the sentence being properly understood - such a doubt is "not a doubt about the truth of an affirmation, but rather an uncertainty about what assertion my sentence represents – whether the rules in virtue of which the sentence means an affirmation to me are the same as those whereby myself and others have previously employed the symbol 'yellow'," (Schlick, 1935b, [1979b, p.411]). But on the Davisonian account we are considering, we can't really prise those ideas apart – what the sentence means is what it is properly understood to mean by the linguistic community, and using language consistently with how we have always used it in our community is, we have argued, essential to communication.

The implication of this is that as speakers we can never be certain that a sentence which we are uttering of the form "here, now, yellow," is true. We have to believe that when we utter that sentence we are doing so with the usual meanings of each of those words, but we do not have guaranteed access to what the usual meaning of those words *is*. As interpreters, the problem is just as pronounced. We assume that the speaker is using words in the same way as they and we have always done, but we don't have any more access to that than the speakers do. Schlick's account is one on which the rules of use for words are chosen by the speaker, and if we allow the speaker to do this on each occasion of utterance then it may be true that they can form sentences about immediate experience which are not open to questioning, but this is to abandon the principle that language use should be consistent.

From these considerations it must be concluded that affirmations, as statements, are not a viable foundation upon which to build the system of science. However, there is reason to be optimistic that we can take the system of science as a whole to be at least mostly-accurate. When the subject matter of our assertions is whatever structures of experience generally cause us to assert those things, our overall network of beliefs is guaranteed to be at least mostly-accurate. In the next section I will develop this idea in more detail, building on the recent work of David Chalmers on philosophical construction.

6.4 Chalmers' Scrutability Framework

Chalmers' project is an investigation into the kinds of philosophical construction that are possible, of which Carnap's work in the *Aufbau* is just one example, which is to say the extent to which it is possible to reduce our various kinds of claims about the world to a compact class of claims about relatively few basic entities. In this section I will use this framework to explore the kind of view that we're left with once we take onboard the philosophical progress that has been made since Schlick was writing. I'll argue that there is room for a view which, although different from Schlick's, at least retains the spirit of his emphasis on empiricism and the epistemology of science.

To be clear from the outset, Chalmers' project is not equivalent to Schlick's – they do not have the same kinds of objective – but Chalmers does provide a framework *within which* a project in the spirit of Schlick's can be investigated. Chalmers' project is a metaphysical one, which looks to see what the minimal class of truths is from which other truths would be knowable in principle. It does not imply that the minimal class is itself knowable – if we need something in order to derive the rest of the truths then we can just add that to the class without worrying about how we might learn it – and the knowable-in-principle relation can be idealised to a godlike intellect without compromising the account. None of this is will get us very far with a project that concerns itself with the actual practice of

science. What we do get from the framework is a clear relation between the core class of sentences and the rest of the system, so, if we can independently establish that some sentences in the core are knowable, then we have already got the relationship of those sentences to the rest of the system.

The total class of true statements is vast, and it includes a number of statements which we might *prima facie* worry it is difficult to reduce to mere physical or phenomenal statements. There are, for example, statements about nebulous entities, like "Britain is under a Conservative government", indexical statements, like "I am in Scotland", moral statements, like "it's wrong to kill people for fun", and mathematical statements, like "the Gödel sentence is true in Peano arithmetic". Chalmers' project is to investigate the various ways we might try to reduce these and other statements to others in a compact class. Chalmers' approach is very general and tries not to take stances on contentious philosophical questions, like the nature of propositions or the comprehensibility of metaphysical questions. Since the position we have been exploring *does not* stay silent on these questions, we will only need to develop a relatively restricted view within the scrutability framework.

6.4.1 Scrutability

In the latter half of the 20th century, a number of philosophers produced cases in which it seemed like the meaning of a term was not determined by the beliefs of the speaker. Saul Kripke (1980) argued that the meaning of a name such as "Gödel" could not be equivalent to a definite description such as "the man who proved the incompleteness of arithmetic", even for someone who knew nothing else than that about Gödel, because we can imagine situations in which the description is wrong (e.g. Gödel stole the proof from someone else) and "Gödel" would still refer to Gödel. Hilary Putnam (1973, [2008]) argued that "water" refers to a substance that is necessarily H_2O, and it did so even before we had an understanding of chemical structures, so that an indiscernibly different substance would nevertheless not have been rightly called "water". And Tyler Burge (1979) argued that two people in different linguistic communities could live their lives with identical experiences and beliefs and yet have their words mean different things, for example if in one

community “arthritis” means what it does in English but in the other it applies to a variety of conditions then the first man would be mistaken in saying “I have arthritis in my thigh” where the other man could present the same symptoms and be correct. What these examples have been taken to show, if we accept the intuitions, is that the meaning of words depends at least in part on factors external to the speaker – things which the speaker might not have access to – and therefore meaning is not necessarily transparent to competent language users.

In light of these considerations, Chalmers does not use anything as strong as sameness-of-meaning or definition to ground his construction. Instead, he uses the relation of scrutability to connect the sentences in the compact class to the class of all truths. Where we have previously said that certain patches of green and brown are the meaning of “there is a tree”, Chalmers’ scrutability relation would merely claim that from knowledge of certain green and brown patches an ideal reasoner would be in a position to know the truth value of “there is a tree”. Formally, there are a few different kinds of scrutability, but Chalmers shows that they are closely related enough that we only need to concern ourselves with one:

> ***A Priori* Scrutability:** S is *a priori* scrutable from C for s [at t in w] iff [at t in w] s is in a position to know *a priori* that if C, then S.
> - Chalmers, 2012, p.40

…where S is a sentence in s’s language and C is a compact class of sentences, again in s’s language (and we stipulate that “if C then…” means “if *conjunction of sentences in C* then…”). To know a class of sentences is to know the conjunction of all the sentences in the class, so we can set aside worries about closure. The phrase “in a position to know” encodes a heavy amount of idealisation, and this will be important as we distinguish Chalmers’ view from Schlick’s. It might not be the case that “if C, then S” is *actually* knowable for any human intellect, but we suppose that a godlike intellect which never makes a misstep in reasoning and appropriately considers all relevant evidence, given an arbitrarily long time to consider the question, would, if they cared to consider it, come to the conclusion that S (Chalmers, 2012, pp.62-63).

The notion of scrutability is not subject to counterexamples in the same way as a definition-based construction. Any purported counterexample will proceed by describing a case and then relying on the fact that from the case described we know what the truth value of the sentence in question is (and that it is not as the proposed definition predicts). But knowing the truth value of the sentence from the description of the case is just what it is for a sentence to be scrutable from a compact base. In Kripke's Gödel case, for example, we are told that the speaker believes "Gödel" refers to the one who proved the incompleteness of arithmetic, and we are told that the incompleteness theorem was proved by a man commonly referred to by "Schmidt" rather than the man commonly referred to by "Gödel", and we are invited to have the intuition that when the speaker says "Gödel" he nevertheless refers to the man whom most people call "Gödel" instead of the one who proved the incompleteness of arithmetic. This might be taken to show that the meaning of "Gödel" is not "the person who proved the incompleteness of arithmetic", but the example does not, and indeed *cannot* show that a person with sufficient information about the world does not know who "Gödel" refers to, because *we* are the people with sufficient information about the world and our intuitions will always be that "Gödel" refers to whoever we believe "Gödel" refers to in the described case. Therefore, the notion of scrutability is safe from externalist counterexample. It is always the case that if we have a complete description of the world then we know what our words refer to.

This immunity to counterexample comes as a result of Chalmers' high degree of flexibility about what can be put in the base. In the description of a putative counterexample we can specify whatever we want about the world, including, if necessary, the brute fact "'Gödel' refers to Gödel and not Schmidt". This may not carry over to a system which takes a principled base for the purposes of, say, constructing the world epistemologically, which is something we must bear in mind as we progress. If we are laying the groundwork for an epistemological construction within the broader scrutability framework, then we will ultimately need to ensure that the base is compressed to a class of sentences that can themselves be known, otherwise there will be unknowable truths in the system.

As a terminological aside, Chalmers cashes out the scrutability theses in terms of sentences rather than propositions. This might be problematic if we think of sentences as uninterpreted sequences of signs, but in Chalmers' account the sentences are uttered by people with particular thoughts in mind. Chalmers is using sentences rather than propositions because there is debate about the correct view of propositions, while sentences are relatively simple. Sentences are clearly *related* to propositions, and there's not too much difficulty in using this relation to set up something we might as well call "scrutability" between sentences, even though the only thing we can really be in a position to know is a proposition. We can say that a sentence outside the base is scrutable from sentences inside the base if the outside sentence expresses a proposition that is scrutable from the propositions expressed by the sentences inside the base. Chalmers does it differently, using an account of thoughts and an *Expression* relation that is different to the one we've been using in previous chapters (Chalmers, 2012, pp.72-77), but he only does so because he has to satisfy people using possible worlds semantics or Russellian propositions, and I have no such concerns here. Chalmers' use of sentences rather than propositions has no effect on what we're interested in.

There is, however, a significant difference which is immediately obvious between Chalmers' construction and the views we have attributed to Schlick. Chalmers' constructive relation is knowledge preserving – if you know the base then you (can) know the rest of the system. In contrast, we have seen that for Schlick at least the majority of claims in science are hypotheses, and only statements about immediate experience might possibly be known. The system of science has to be in some sense based on these moments of experience, otherwise it could scarcely claim to be a Schlick-inspired model, but only in a belief-justifying way, not in an absolutely-confirming way. The difference here is a result of the idealisation – Chalmers' base potentially includes all the experiences you will ever have and considers what you can know if you know them all at once, whereas Schlick is talking about what it is actually possible to know – again, Chalmers' construction is a metaphysical one while Schlick's is epistemic. But we have had persistent difficulty in getting infallible propositions to relate to fallible propositions, because

disconfirmation of the fallible proposition would seem to equally disconfirm whatever proposition supported it, and therefore the fallibility must spread throughout the system. If epistemic properties apply to the system as a whole and not to individual propositions, then the epistemic construction we end up with must justify belief in the system of science as a whole rather than just a small class of epistemically foundational propositions, and we will conclude that such a justification is possible in section 6.4.3. Insofar as affirmations are supposed to hold a privileged epistemic status over the system of science, then, we should be clear that in what follows we will consider affirmations to stand *outside* the system. Given that the scrutability thesis says that all truths are knowable from the compact class, this suggests that we may not be able to interpret affirmations as truths at all in this framework. We will come back to that issue in section 6.4.3. In the system we eventually come to, we will consider protocol sentences as a scrutability base and consider their epistemic properties as separate from those of affirmations. We will ultimately be looking to justify acceptance of protocol sentences, but this justification may be weaker than the absolute infallibilism that is required of affirmations.

6.4.2 The Scrutability Base

Scrutability is always scrutability *from* something, which we hope will be a class of statements that is as compact as possible. ("Someone is in a position to know all truths if they already know pretty much all truths," isn't an interesting thesis.) Chalmers spends most of his time talking about scrutability from base PQTI, which is a unity of four classes of truths. P is the class of physical truths, Q is the class of phenomenal truths, T is what he calls a "that's-all truth", and I is the class of indexical truths (Chalmers, 2012, p.110ff).

P – Physical Truths include both microphysical and macrophysical truths. They will be expressed in some complete language of physics and say things about particles, spaciotemporal properties, mass, electromagnetics, and so on, in addition to sentences about large composite objects. It is immediately obvious that there will be some overlap between microphysical and macrophysical truths and we may not need both, but this will be addressed below when we start compressing the base.

Q – Phenomenal Truths are what it is like to be a creature at a particular time. Separately to the scrutability project, Chalmers is famously committed to the view that it is possible for people to exist who act indiscernibly from ordinary people but have no conscious experience ("philosophical zombies"). He also believes that, in Frank Jackson's Mary example (see pages 76-77), Mary learns something new upon leaving the room and experiencing red for the first time. That means that phenomenal truths, for Chalmers, will include sentences which purport to express content. We have argued that no such sentences can be meaningful, but even if we take those out of the base there are still other phenomenal truths in this class. The relation of perceptual similarity which Carnap uses in the *Aufbau* is included in this category, and so more generally will be sentences about the structure of experience. We have argued before that sentences about the structure of experience are not a separate category from sentences about the real world – again, this will be addressed below when we start compressing the base.

T – The "That's All" Truth is designed to avoid the worry that a class of true statements doesn't rule out other unrelated things from being true. A complete description of the physical world, even with the understanding that it really is a complete description, isn't enough to say that there isn't a non-physical ectoplasmic reality as well. Chalmers cashes T out as a claim that the world is a minimal scenario required to accommodate all of PQI (p.111), and I won't oppose this here.

I – Indexical Truths are the truths that locate the speaker at a particular place and time. They allow us to have sentences such as "the cup is to my right", which are true but not derivable from physical/phenomenal truths alone.

PQTI is clearly a much larger base than Carnap (or Schlick) used (in the 1920's and 30's), but there are a lot of truths that are not in these categories but are nonetheless scrutable from them. For example, moral truths are taken to be knowable from this class if they exist at all. The belief that there are moral truths stems from the fact that in many describable situations we know which actions are right and which are wrong. We may not be able to articulate exactly what features of the case are relevant to this decision, but that doesn't matter for Chalmers because

he isn't reducing moral truths by way of definition – just by scrutability. In the real world there often appear to be times when people disagree over which course of action is right, or are simply not sure, but these cases can often be put down to disagreement over empirical facts or less-than-ideal reasoning. If disagreement were to remain between ideal reasoners in cases where the empirical facts were plain to all, then this would appear to be a case in which there really was no moral truth of the matter, and in that case the scrutability thesis remains unthreatened. Moral truths are just one kind of truths which are left out of the base but taken to be derivable from it, but the method is similar for others. Ultimately, there are very few claims which people hold true that cannot be known by someone whose knowledge base includes PQTI, and those that exist are positions like epistemicism about vagueness – the fact that they are not scrutable from PQTI, even for a godlike intellect, is a fairly good reason to hold that they are not truths.

Ultimately, Chalmers thinks that the base can be reduced much further than just PQTI, but not as far as Carnap's project which uses just one phenomenal relation and a few logical expressions for everything. For example, it's fairly easy to reduce macrophysical truths out of the base by looking at how they will be scrutable from microphysical truths. If you know the location of every atom in the Eifel Tower, then you're in a position to know the location of the Eifel Tower; if you know the mass of every atom in the Eifel Tower, then you're in a position to know the mass of the Eifel Tower; and so on. Carnap's original reductions will do a lot of work for us here, even if the last step has to be rethought. We can reduce everything that can be said about behaviourist psychology (allowing for the moment that there might be some non-behaviourist psychology) to macrophysical truths about the behaviour of bodies in response to certain situations. We can reduce social customs to behaviourist psychology. Really, it looks as though everything that can be known by science can be compressed into statements about the microphysical.

Chalmers has his own preferred base, but highlights four key points of choice that, depending on your own philosophical commitments, will be consistent with the scrutability thesis (Chalmers, 2012, p.357ff). For example, if you think that there

is a difference between possible worlds with natural laws that are never in fact instantiated and physically-identical worlds which don't have those laws, then you need nomic truths in the base. We can imagine a world that physically consists of only a single atom, and so there is never any gravitational attraction between objects. Can we describe this world in two ways, both with and without gravity? Or is gravity an observed regularity between objects, and therefore necessarily not a feature of a one-object world? Chalmers thinks we can make sense of the distinction, and so includes nomic truths in his base (2012, pp.338-340). Schlick, when discussing how natural laws differ from definitions in maths, says "each of the symbols used [in the expression of a natural law] refers to definite observations and experiments which actually must be carried out, if their statements are to have any meaning whatsoever," (1935d, [1979b, p.441]), so for him the one-atom world does not have a law of gravity. So, on our account, while it might minimise the base to reduce things to physical laws rather than a description of all physical objects at every possible time interval, what's important is that we don't need both. Similarly, a Schlick/Carnap-inspired base will take the negative choice at each of Chalmers' other points: spacetime concepts are functionalist, not primitive (i.e. there's no absolute distance, things are just bigger or smaller than one another); our position on phenomenal truths is antirealism rather than realism (i.e. there is nothing that can be known about content, although there are facts about the *structure* of experience); and quiddities (claims about the intrinsic properties of things beyond their relations to other things) are nonsense.

The account that we're building here is clearly different than that which Schlick was putting forward, not just in method but in scope. Chalmers' project takes it as a starting point that the statements we make in everyday discourse are meaningful, and therefore tries to accommodate the fact that they might not all be connected to the language of physics. He thinks that we might need several completely distinct groups of statement to be able to derive all truths from them, in contrast to the early positivist view that it was possible to unify the languages of science so that all that could be said could be said in the language of physics.

Furthermore, Chalmers is looking for a way for all truths to be knowable from a compact class, where we have seen that Schlick only argued that not everything was doubtable. Indeed, Schlick thought that the majority of scientific claims could never be finally verified and always remained open to revision. We said above that this difference came as a result of the idealisation that Chalmers brings to the picture. To follow through Chalmers' arguments we are to suppose that we know everything that we will ever perceive all at once (though the conclusions of these arguments are of course phrased as the conditionals "*if* someone knew the truths in the base, then…"). No doubt Schlick would agree that if we could know all the protocol sentences that would ever be true at one time, then the system of science as a whole would be immune to disconfirmation, since we would already have knowledge of any evidence that might have disconfirmed it. But this highlights the difference in the scope of the two projects, because Schlick is looking for epistemic certainty that we might actually obtain in the course of scientific research, and is prepared to search outside of the system itself for it, whereas Chalmers is interested in the question of whether there is anything beyond propositions about experience that we might need in our base for the derivation of all truths.

Nevertheless, if Chalmers produces a workable account in this vein then it will help us produce a workable model of epistemic security for the system of science. What Chalmers gives us in the first instance, and keeps clearly in view throughout, is a clear connection between the propositions in the base and the propositions in the overall network of truths: a priori scrutability. And as I have said, what we are most interested in here is the *framework* within which we can make some changes to investigate a Schlick-inspired position. As noted, Chalmers highlights a number of points as he develops his position at which other people would make different choices depending on prior commitments, and of particular interest will be his account of how Carnap's construction would look if it were to deal with the problems that have since been highlighted for structuralist accounts. So, let's look at what happens if we take the Schlick option at every point of choice.

For reasons of space it isn't possible here to give a complete, detailed account of the way a Schlick-inspired base will be compressed, but we can give an outline of how the project would proceed. Our base will start with a Ramsified sentence that describes the objective features of the world in logical vocabulary. A big question here is whether this sentence will be describing the structure of the physical world or of phenomenal experience. We are looking for a view with a unified language of science, but we have seen that Chalmers has separated the two in the base (at least, before compression). We have argued before that there is no meaningful distinction here for Schlick (see e.g. section 3.1.3), and claims like "the post box looks red" say no more than claims like "the post box reflects the same wavelength of light as other red objects". All I will say here is that we will continue with this approach within the scrutability framework, so P and Q are considered equally reducible to some structural fact about the world. Chalmers' considerations about structuralism are important for securing the epistemic grounds for the system as a whole, and so we will go through this in more detail in the next section.

There still needs to be a "that's all" sentence, which requires that the notion of fundamentality appears in the base – we need to be able to say that there are no additional objects than the minimum necessary to satisfy the claims in the base. At first sight it might look like it should be possible to put "that's all" into the construction relation, since it seems like a merely procedural requirement that we not end up with any more fundamental objects in the conclusions of our reasoning than were present in the premises, but the "that's all" truth is remarkably resistant to this approach. We would be changing the *a priori* scrutability thesis from "for all true statements S, a subject who knows the base C is in a position to know *a priori* that S" to something like "for all S, a subject who knows C *and knows that there are no fundamental objects besides those necessary to instantiate C* is in a position to know *a priori* that S", and that's just the same as adding a new sentence to C.

With regard to I, the Schlick/Carnap approach, continuing to set aside affirmations for the moment, is to say that we're only interested in impersonal claims, so that indexical statements don't appear in the set of truths which we want

to construct. That allows us to do without I. So, our extreme logical reduction ends up with a single Ramsey sentence and the claim that that's all there is.

The biggest problem with a Carnap-style structuralist reduction is the objection raised by Newman against Russell in 1928 (but not fully recognised by Carnap until later) – any model is satisfiable by any set of the right cardinality. (This is the objection to which we alluded in section 3.3.5 and looked at under the name "the foundedness problem" on page 95.) This is because, if we are sufficiently relaxed about what counts as a relation, we can satisfy the model by substituting for each relation a set of ordered pairs of objects between which the relation is said to hold. Carnap's structure descriptions rely on there being a limited number of relations in the world for the model to map onto, but as far as logic is concerned there are possible relations between any two objects that we can use. Relations can just be defined arbitrarily as sets of object pairs. Carnap attempted to mitigate this worry by specifying that we only map the structure onto "experienceable 'natural' relations," (1928, [1967, §154]), but Chalmers points out that this is to introduce new terms into the vocabulary and therefore new truths into the base (2012, p.9). So, a model that only describes logical relations without naming the things that are being related says nothing about the world except that there are a certain number of things in it, but it is possible to avoid Newman's problem if we supplement the logical construction with something else. Well, we already have something else in the base – T – and Chalmers points out that the notion of fundamentality that it brings in can be given an expanded role.

Carnap's aim was a purely structural account, but Chalmers offers a position he calls "weak structuralism" that comes close (2012, p.409). On the weak structuralist account, the logical vocabulary is expanded with expressions for a limited number of relations. Carnap's original relation of phenomenal similarity would be one option here, preserving the greater part of his construction. Our scrutability thesis would now be "all truths are scrutable from truths using logical vocabulary plus structural expressions," (Chalmers, 2012, p.409). "Structural expressions" here can be interpreted in a variety of ways. If we make them

spatiotemporal properties, for example (i.e. we say something like "this model maps onto reality and (to avoid vacuity) these three relations are spatial and this one temporal"), then we very quickly get to a physicalist account, although this would require that there actually be some matter of fact about relations of space and time, which is broadly unacceptable for reasons covered above. Moreover, the goal of this section is to show how Schlick's emphasis on experience can be incorporated into a viable framework. We need P to be derived from experience rather than the other way around.

A base more conducive to our aims is that of weak *phenomenal* structuralism. On this view, all truths are scrutable from phenomenal truths and phenomenal truths are taken to be structural in nature. This is the ontological account that runs closest to the epistemological one we have been expositing for Schlick (and if we believe that all possible truths are knowable in principle, as Schlick did, then it also appears to be the metaphysical construction that Schlick would have to endorse (insofar as he would consider the question meaningful at all)), and Chalmers helpfully explains what this will mean within the scrutability framework. In order to escape worries that Carnap's phenomenal-similarity relation cannot fully capture the varieties of experience, Chalmers suggests replacing it with a parametric relation which captures information along multiple phenomenal dimensions, giving, for example, visual perceptual data of an individual at a time as a function from locations in their visual field to colours (*a*, *b*, *c*) corresponding to positions along the red, blue, and green colour axes (2012, pp.413-414). This account ties in easily with the position we have articulated for Schlick in the past five chapters. It does retain the problem that we need a primitive notion of a phenomenal relation in our ontological construction for our structure description to map onto. But if we accept this then our epistemic construction becomes much more plausible than if we were using a primitive spacetime concept, because *relation that you experience* is something we might plausibly be able to learn about from experience. Moreover, the account continues to map easily onto Schlick's philosophy, because we have seen (e.g. section 3.1.2) that Schlick thinks we can't get away from the requirement that languages be linked to experience. Necessarily, the structural relations that we learn

to associate with words are forms of experience – relations that we are able to identify between colours, sounds, textures, and so on – and this means that the class of true propositions will necessarily be expressed in a phenomenal language, since no other language is possible. Schlick's position never falls into the Newman objection because Schlick never thought we could adopt a language without reference to experience. Chalmers has a residual objection that this account doesn't distinguish between the different ways in which something can be experienced, allowing that colours for one person might be experienced like sounds to another (2012, p.415), but this was an objection we dispensed with back in chapter 3 (section 3.3.2). (He also brings up Mary.)

Chalmers does mention one other position that we can't rule out immediately, which he calls "fundamentality structuralism".[34] On this view, rather than specifying that some particular relations in the Ramsey sentence pick out phenomenal relations, we specify that they pick out *fundamental* relations, thus making use of the extra notion that we already had to introduce to state the "that's all" truth. We are required to have some primitive understanding of what it is to be fundamental, but again, we needed that anyway, and Chalmers argues that "it's not obvious why this is any more problematic than a primitive grasp of the notions of conjunction and existential quantification," (2012, p.418). Chalmers also claims that this view is most closely related to the position which Carnap seems to be taking with his recourse to "experienceable 'natural' relations", conveniently dropping the "experienceable" part of the quote at this stage in the argument. The fact that Carnap put "experienceable" and "natural" next to each other so readily highlights why I'm not concerned about this alternative – on a certain view about what "fundamental" means, fundamentality structuralism just collapses back into phenomenal structuralism.

[34] He also mentions two positions – nomic structuralism and quiddistic structuralism – that we *can* rule out immediately for obvious reasons.

The overall framework we are now adopting looks like this: All true propositions can, in principle, be known if a subject knows all of the truths about the structure of experience across time and location. The total network of true propositions is entirely structural, which is to say that we can replace all of the names for objects and relations with variables and not lose any information. If we have the Ramsified network of propositions and investigate the world for experienceable relations we will find, given enough empirical investigation, that there is a unique assignment of experienceable relations to variables in the network that results in the experienced facts corresponding with the propositions in the network. Protocol sentences are individual propositions about the phenomenal structure of the world from a certain perspective at a certain time, and they have a logical place in the network of truths based on the fact that knowing all of them would be sufficient to know all of the truths, although we have not at this point argued that it is ever possible to know any of them. In the next section, however, we will see that there is good reason to think that the majority of the protocol sentences that we accept will be true.

6.4.3 Foundationalist Epistemology in the Framework

Having re-established a coherent structuralist position along the lines of the *Aufbau*, is there room to establish secure epistemic foundations for the system? The scrutability thesis is that *if* we know the base, then we know the rest of the truths in the system, but getting to know the base is the tricky part. The problem we have had, in-play in later Wittgenstein and highlighted by Davidson, is getting the foundational statements to be infallible once they are tied to the system of science. We have seen that if they are to be coherent sentences their meaning must be fixed, and we have adopted from Chalmers the view that for this to be the case is to agree in the same circumstances that they are true. Furthermore, a meaningful statement, as we have seen time and again, cannot be entirely demonstrative. Meaning cannot be fixed as "whatever is currently true" – in the intensional framework we are now considering, we must be prepared to accept on any occasion of use that in a different possible world we would regard the same proposition as false. That is, any infallibility must be epistemic and not logical (as we said on pages 140-142).

The Schlick-analogue to Chalmers' project, if it's going to be in the spirit of Schlick's philosophy, is firstly going to have to give a central role to experience and secondly attempt to use that experience as a catalyst for knowledge. We laid the groundwork for experience to have a central role by insisting on phenomenal structuralism rather than mere fundamentality structuralism. Our construction doesn't take the phenomenal base as a mere conventional choice – it takes it because the process of language-learning seems to guarantee that terms will be associated with structures of experience, meaning that the relations quantified over in a Ramsified protocol sentence must be relations that are found in experience, not arbitrary object-pairs and not objective spaciotemporal relations which it's not clear we can even make sense of. But these protocol sentences don't have the epistemic properties of affirmations that we investigated in the last chapter. Critically, protocol sentences are formed in a unified, consistent language, and so it is possible to make mistakes about whether what we are currently experiencing corresponds to a given protocol sentence. We don't get to fix the rules of interpretation just based on what we believe to be the case in the moment of utterance.

Individual protocol sentences must be fallible, but we saw from Davidson (section 6.1.3) that it might be possible to consider our network of beliefs to be justified as a whole if we can guard against being systematically wrong. Davidson's view was that meaning is related to whatever systematically causes utterances, so systematic deception is impossible, but we argued that meaning had to be based in structures of experience rather than inaccessible causes. Nevertheless, it might be possible to take a system-wide approach to justification rather than a statement-by-statement approach, and Chalmers develops some thoughts on the subject (in line with the *Aufbau* rather than the work which came after it). Chalmers construes Carnap's anti-sceptical argument as follows:

1. Ordinary beliefs have only structural contents.
2. The structural contents of most ordinary beliefs are justified.

   Therefore:

3. Most of our ordinary beliefs are justified.

- Chalmers, 2012, p.432

Chalmers proceeds to argue that the force of this argument does not depend upon a phenomenal structuralism, but we don't need to worry about that because we're quite happy to accept a phenomenal structuralism, and he certainly doesn't argue that it depends on a *rejection* of phenomenal structuralism.

The argument is, in the same vein as Davidson's anti-sceptical position, an argument against global scepticism rather than scepticism about any individual claim. Global-sceptical worries are usually based on scenarios that we cannot rule out in which an important class of our beliefs are caused by something that, intuitively, is not what we believe causes our beliefs. Our account, however, has been that "things in themselves", so to speak, don't appear in statements, either of the protocol base or the set of truths derived from it. Our Ramsified account of the physical world, for instance, doesn't say anything about *what* it is that plays the electron role, so that there is no difference between our world and a world in which quasi-electrons, identical in every respect to electrons but different in "quiddity", play the role. Similarly, in any grand sceptical scenario such as Descartes' evil demon, there will be *something* playing the electron role. The illusion has features that correspond to what it is trying to simulate in such a way that we can map the same structure description onto the illusion as we can the real world. So, where the sceptical hypothesis claims that we can't know that such-and-such illusory scenario doesn't obtain and in that scenario most of our beliefs are false, the structuralist response accepts that we can't know the scenario doesn't obtain but says that most of our beliefs are still true in that scenario.

We saw in the Davidsonian account (section 6.1.3) that his response to global scepticism, as opposed to local scepticism, was designed to sidestep Schlick's requirement that there be certain foundational truths that provide an epistemic guarantee to the rest of the system of science. He makes the claim that the things that ordinarily justify our beliefs can't be mistaken in a systematic way that throws off the whole system, while still allowing that any individual belief can be mistaken

in a variety of non-systematic ways. The same overall result comes out of Chalmers' argument, which allows that we might be wrong locally about the structure of the world even if it doesn't matter what is instantiating that structure. Grand sceptical scenarios are ruled out, but mistakes in specific instances remain possible. But the scrutability framework is fundamentally foundationalist in nature – it builds in a base of truths from which the rest are knowable. Our worry, then, should be specifically about the truths in the base rather than the "quiddity" of the objects which instantiate the structure.

Securing the base independently of the security of the overall system requires something beyond what Chalmers has given us. As noted, the knowledge-preserving property of the scrutability relation means that any epistemic properties of the base are going to be shared across the system, which rules out having a special epistemic status for protocol statements in particular. We have also noted that it seems like it will be impossible as a matter of fact to know all of the propositions in the base. Schlick's contention has always been that we can know only those propositions that describe first-personal immediate experience – affirmations – which so far we have not been able to bring into the system at all.

In section 4.4.4 we saw how Neurath intended to give a special role in science to protocol sentences, despite his having recognised their lack of epistemic privilege. For him, they guaranteed a link to experience by a stipulated structure that required a link to perception and ensured that the only genuine protocol sentences were those that were actually asserted. Adoption of that structure for protocol sentences was a pragmatic matter, making explicit the features of the system of science that Neurath wanted to emphasise, but the features were a matter of fact: Neurath's conception of science is as a social endeavour based around reports of experience, which gives those experience reports a special role even without a special epistemic status. Schlick's project had a different objective to Neurath's. He wasn't trying to describe the process of science; he was trying to find an epistemically secure foundation for scientific knowledge. It's an idealistic project rather than a pragmatic description of a social phenomenon, so it isn't appropriate to rely on the actual

practices of assertion like Neurath did – if we can imagine a world in which scientists assert internally consistent fabrications, that's a problem. We have seen that Schlick attempted to ground science in affirmations – first-person statements about moments of experience which were supposed to be infallible. We have subsequently argued that the base of a system of science can't have a stronger epistemic status than the system as a whole, and so if affirmations are infallible then they cannot have a deductive link to the system, but there nevertheless appears to be a role for moments of experience. Because, if protocol sentences have no more security than any other sentence, why should we consider them to be the base from which the rest of the system is scrutable, rather than considering the protocol sentences to be a mere consequence of higher-level scientific truths? A perfectly reasonable answer seems to be that they are the kind of thing that we can justifiably come to believe on the basis of moments of experience.

We have said that all truths should be scrutable from a base of weak phenomenal structuralism, but we have also said that we only have access to a small part of this base. This limited access is the reason why the majority of claims in science can never be more than hypotheses – we are always at risk of finding disconfirming evidence in the future. We have said that the protocol sentences that we adopt cannot all be false, because what they assert is in part determined by the structures that systematically cause them. This gives the protocol sentences that we adopt a kind of *prima facie* security. But why do we adopt these protocol sentences? It can only be on the basis of moments of experience. We have not managed to build infallible affirmations into the system, but the protocol sentences that we adopt moment by moment are clearly adopted on the grounds that we believe that they describe the experiences we are currently having, and since they will, necessarily, usually describe the experiences that we are having at the time when we utter them, first-person present-tense experience statements seem to form the epistemic foundations of our actual network of beliefs, not individually, but as a group over time.

Affirmations, on this account, cannot be *truths*, per se, because if they were then they would have to be derivable from the scrutability base. Affirmations are the sentences we are prompted to assert in response to our experience, and the fact that we are prompted to assert them confers justification on a corresponding protocol sentence. They cannot be wrong, because it is never false that some experience prompted them, but the protocol sentence that they justify might have to be revised if it cannot be made to cohere with other accepted protocol sentences, and moreover this divorce from the overall system means that they cannot really be right either. They play the role of connecting the protocol sentences of an individual to that individual's experience, but don't really say anything in the process. It would still be fair to describe them as Schlick did, as the "points of contact between knowledge and reality" (UFE, p.387) and as "moments of fulfilment" (UFE, p.387), but we have not been able to give them a stronger epistemic status than the system of science as a whole. I think that this is as close as we can get to Schlick's view within a consistent construction, but we must acknowledge that it is not the view that Schlick was trying to put forward in "The Foundations of Knowledge".

The affirmations of others can play a further role in expanding our knowledge base by justifying the adoption of protocol sentences about other peoples' experiences in much the same way as we considered with the criterial relation in chapter 5. There, our objection was that we have some independent conditions for the truth of a protocol statement than just the assertions of the individual concerned, and it is easy to construct cases in which one person's affirmation would not justify the acceptance of the corresponding protocol sentence (e.g. if they are under the influence of hallucinogenic drugs). However, it is equally clear that when affirmations and our own experiences are considered as a group, we have no further evidence to draw on. The single affirmation "here, now, blue," may not be the criterion for the acceptance of the protocol statement "S saw blue on Thursday", but some combination of affirmations collectively seem like they can fill the role of criteria, because it is inconceivable that we might come across better evidence than what is given to us in experience. Therefore, a base of protocol sentences can be justified by recourse to affirmations which, altogether, form the

criteria for the acceptance of the base. Those protocol sentences that we are justified in accepting will be only a small part of the total base required from which we might derive all truths, but their relationship to the system of science as a whole is perfectly clear.

6.5 Conclusion

We have seen that Schlick's objective was, from the start, to establish firm epistemic foundations for science, and we have looked at how he tried to do this in the only way he could see to ensure that those foundations were truly empirical. Schlick attempted to show that sentences about immediate first-person experience were infallible in the moment of utterance, and thus could be used to confirm or disconfirm the predictions of the system of science. Neurath argued that any sentence could be disconfirmed in principle, it being a defining feature of sentences that they can be true or false, and that there was nothing epistemically special about affirmations or protocol sentences that should lead us to hold them true over all others. He argued that it was impossible to compare sentences with reality because when we seek the justification of a sentence it is impossible to appeal to anything other than another sentence. Nevertheless, we saw from Schlick's response to Neurath and from Davidson's later work that the only possible understanding of "true" is a correspondence theory, rather than the coherence theory which is often attributed (fairly or not) to Neurath. A sentence is true iff it corresponds to a fact and that fact actually obtains in the world.

We saw in chapters 2 and 3 the way in which Schlick attempted to flesh out this correspondence relation along the same lines as Wittgenstein had in the *Tractatus*. On Schlick's account, true sentences have a logical form that matches the logical form of objects arranged in the world. Forming an affirmation is then a question of reproducing the logical form of experience in a proposition, and in chapter 5 we looked at a few ways in which we might attempt this. However, it does appear that any sentence which maintains its logical relationships with the rest of

the system of science will fail to exhibit the infallibility we are looking for. Either the sentence says something substantive, in which case it can be mistaken, or it says some variation of "here now this" which is empty. Schlick might have thought that the philosophy of language emerging from Wittgenstein would bridge this gap, by speakers' private-language affirmations fulfilling the criteria of intersubjective protocol sentences. This would allow affirmations to be infallible because they wouldn't bear deductively on the system of science, but they would nevertheless bear on the justification of scientific claims. However, this account did not appear to be viable for propositions with truth conditions that we have independent access to, such as "S sees blue" (in contrast to "S is in pain").

Davidson gives us a different kind of correspondence relation between sentences and the world, on which a sentence corresponds to whatever fact causes us to utter it in normal circumstances. This gives us an account of truth on which it is impossible for the majority of a person's everyday beliefs to be false. We saw that an account like this would still be heavily reliant on experience, and indeed that Davidson's argument proceeds in much the same way as Schlick's when it comes to the question of how we learn a language. We also saw, however, that this introduces a further problem for Schlick's account. Schlick's conception of grammar is one on which individuals use language in accordance with rules that determine when it is appropriate to make certain utterances, and these rules are selected by the speaker. On Davidson's account, however, speakers are not necessarily aware of the rules that underlie their communication. This means that there is no hope of reproducing the form of experience in a sentence in any way that would guarantee success. If speakers are not necessarily aware of the rules governing the interpretation of their sentences, then they cannot use those rules to reliably translate experience into words. The best we can do is use the rules which we *believe* underlie language. This means that any sentence within the system of science is epistemically fallible, implying that no protocol sentence can ever be absolutely confirmed.

We have attempted to salvage some important aspects of Schlick's philosophy and shape them into a philosophically-viable position. If we take

phenomenal structures over time as a base from which all truths are scrutable then we have an ontological construction that provides a clear relationship between protocol sentences and the system of science as a whole. No one person can know the entire class of protocol sentences by themselves, but the epistemological foundations of the system are nevertheless in that class – protocol sentences provide the base from which the rest of the system will be known, and not the other way around. Why is this, when protocol sentences have no stronger epistemic status than the rest of the system? It is because protocol sentences are the kinds of proposition that can be justified on the basis of individual moments of experience. We know that the meanings of protocol sentences are determined by the experiences that systematically prompt assent to them, which means that as a group they cannot be systematically wrong, and this means that affirmations provide justification for the foundations of the system. The foundations are built on soft ground – any individual claim alone cannot hold the weight, and we have not been able to give even affirmations the infallible character Schlick sought after – but the structure as a whole is sturdy and admits no possibility of generalised sceptical hypotheses. If there is something that systematically causes our phenomenal beliefs, that something has enough of a structure for us to investigate it, make claims about it, and predict future experience from it. Thus, Schlick's foremost objective is achieved: the overall system of science is secure.

Overall Conclusion

Where have we ended up? At the outset I said that there were two projects running side-by-side in this thesis, and it will be useful to consider them separately now.

First, we wanted an account of what Schlick's position actually was, with regard to the questions in which we are interested. We have seen that structuralism was a feature of Schlick's philosophy since before he came under the influence of Carnap or Wittgenstein, and it continued to play a role until the end. The motivation for this view was that knowledge always involves knowing *that* something, and correspondingly, a meaningful proposition is always the proposition *that* such-and-such fact obtains. If we bring this reasoning to bear again on Jackson's example of Mary, then it is clear that whatever she gets from leaving the room is something categorically different from the knowledge in which Schlick was interested. No sentence can be formed that is inscrutable for Mary when she is in the room and then learned when she leaves. What would that sentence look like? "When I look at post boxes, I see red"? Mary already knew that. "When I look at post boxes, I see Edenic redness," to borrow a phrase from Chalmers? What does "Edenic redness" mean? If it refers to private sensations, then it has no truth conditions. If it doesn't, it can be known without experience. Maybe there is something that it is like to enjoy red phenomenal content that is separate from the structure of the experience, but if so, then it cannot be *known*. There are no facts about it, so we don't need to worry about it being scrutable from a structural base.

So, every meaningful proposition says that certain relations obtain between certain objects, and says nothing more than that. What is a proposition? For Schlick, we have seen that it was a combination of a physical thing – a sentence – along with the rules of interpretation conferred by the speaker's intentions at the moment of utterance. That which expresses a fact is itself a fact, and in order to express the fact it shares its structure. The structure, after all, is that which is communicated – it is common to the fact, the proposition, and the mental state of the reader/listener. That which is not communicated is what Schlick referred to as Content. *This* is the role that phenomenal redness plays. Something has to instantiate the structure

which is communicated. In the world, the structure is instantiated by objects and physical relations, but in the mind of a speaker/interpreter the structure is instantiated by content. If a rock is between two trees, then in the real world it is the rock and the trees and the relation *x is between y and z* that are quantified over in the Ramsey sentence "∃x ∃y ∃z ∃R (Rxyz)", but when that proposition is grasped by a person looking at the landscape it is instantiated by a collection of phenomenal greens and phenomenal browns and phenomenal rock-colours. When it's written on the page, the structure is instantiated by spatially-arranged marks, in conjunction with the rules of language that tell us what structure particular marks represent.

The rules of language, for Schlick, are arbitrary and chosen by the speaker, and a listener understands a speaker once they know what rules the speaker is using. The only way to coordinate the rules, Schlick thinks, is to indicate to a person who does not yet know them the conditions under which certain claims are true. This, necessarily, leads us to the verification principle. The only rules that can be understood are those that link to situations in which propositions are verified, so the meaning of a proposition is the method of its verification. The verification principle, then, far from being a dogmatic claim by people looking for a way to get rid of metaphysics, is the endpoint of a careful line of inquiry. For Schlick, although the rules of interpretation for a language are arbitrarily chosen, they are so within the restriction that they must link the language to that which can be experienced.

The problem that we have had with Schlick's philosophy arises from the fact that rules are linked with experience and chosen by the speaker. On Schlick's view, a certain class of statements, with rules chosen in a demonstrative manner, are to epistemically underpin the system of science. These statements, affirmations, are such that when uttered sincerely they cannot be false, because the rules for interpreting them are selected so as to link to what the speaker is currently experiencing. Since the rules of interpretation link to the speaker's experience rather than linking to the intersubjective language of science, they cannot deductively imply or confirm anything in the system. Schlick may have thought that affirmations could nevertheless justify acceptance of certain protocol statements in

the same manner as a person crying out after falling justifies the belief that they are in pain without being deductively linked to the claim. But we saw that this approach was problematic for propositions that we can confirm independently, and in the intersubjective system "S sees blue" can be tested in other ways than S's assertions – the claim is true if a certain wavelength of light is detectable before S's eyes, so S's assertion that they see blue is fundamentally of a different kind to their assertion that they are in pain.

Moreover, we had concerns about Schlick's position that speakers could select rules of use in the moment of utterance. The problem with this is that affirmations appear to be phrased in the same language as every other statement, and that tends to bring along with it all sorts of epistemic problems – the belief that these same words have previously been used for this form of experience, or that other people in the community use the words in the same way, for example. The fact that language-use should be consistent is, we saw in chapter 6, just as central to learning it in the first place as any link to experience. Schlick attempts to give an account of affirmations as being not logically but epistemically infallible on the basis of rules of use selected at the moment of formation but nevertheless not equivalent to *whatever is currently the case*, but we can see that he is pulled back towards the intersubjective language by the fact that he continues to use its vocabulary in the formation of affirmations. This seems like it will sacrifice the infallibility of the affirmations.

This brings us on to the second part of the project – the minimum necessary changes to create a viable philosophical position in the spirit of Schlick's endeavour.

Davidson offered an account on which coherence would guarantee a degree of correspondence with reality. On this account, the meanings of terms were to be determined by whatever typically caused our utterances. Then it would be the case that we could be wrong about any individual claim, because the situation in reality might be idiosyncratic with what normally causes similar claims, but we couldn't be wrong across our entire network of beliefs because the thing that we believed was in some sense determined by what was actually the case. We argued that this

account was flawed in that it fails to appreciate the way our internal states clearly play *some* role in meaning, but the approach of rejecting global scepticism whilst accepting some measure of local scepticism was instructive.

A similar approach to rejecting global scepticism can be found in Chalmers' argument from structuralism, but with the important difference that we can get experience into the account and avoid committing ourselves to a coherence-based account of justification. Chalmers' ontological framework gives us a clear relationship between protocol statements and the rest of the system of science, which is that if we know the observation statements then we are in a position to know the rest of the system, rather than that the rest of the system is related to the observation statements by definition. This seems necessary as a response to the externalist challenges of the late 20th century, and it doesn't sacrifice the spirit of the construction. Once this is done, we can argue that no truths are inscrutable from the set of statements about the structure of our experiences at various times. We can allow that individual statements in the set might be idiosyncratic with the rest and so not bear an appropriate relationship to the system of science that we use to explain the experiences, but it is not the case that the entire base can be disconnected from the system. Epistemically, we can know that it is not the case that the majority of our accepted protocol sentences might be false. This is because the beliefs are structural – as Schlick maintained from the beginning, the meaning of a statement is to be found in the connections between the terms employed and not in the denotation of the terms themselves – and, whatever actually caused them, it must be something with an isomorphic structure. Moreover, what causes our adoption of protocol sentences is clearly individual moments of experience. The role of affirmations within our system is as the justification for the acceptance of protocol sentences, and the epistemic importance of protocol sentences is that they are the kinds of proposition that affirmations can justify. The justification is fallible on an individual level, but not across the whole class of accepted protocol sentences, and that means that we can achieve Schlick's long-time objective of grounding the system in statements about experience which, taken together, are epistemically secure.

Bibliography

- Awodey, S., & Carus, A.W., 2001, “Carnap, Completeness, and Categoricity: The ‘Gabelbarkeitssatz’ of 1928”, *Erkenntnis*, 54:2, pp.145-172
- Ayer, A.J., 1936, "Verification and Experience", *Proceedings of the Aristotelian Society* 37, pp.137-156 reprinted in A.J. Ayer (ed.), 1959, *Logical Positivism*, The Free Press, New York, pp.165-198
- Ayer, A.J., 1946, *Language, Truth, and Logic*, reprinted 1952, Dover Publications, New York
- Ayer, A.J., 1959, "Editor's Introduction", *Logical Positivism*, The Free Press inc., New York, pp.3-28
- Baker, G., 2003, “Introduction”, in Wittgenstein & F. Waismann, 2003 *The Voices of Wittgenstein; The Vienna Circle*, G. Baker (ed.), G. Baker, M. Mackert, J. Connolly & V. Politis (trans.), Routledge, London
- Benacerraf, P., 1965, "What Numbers Could Not Be", *Philosophical Review* 74, pp.47-73, reprinted in P. Benacerraf & H. Putnam (eds.), 1983, *Philosophy of Mathematics, Selected Readings*, Cambridge University Press, Cambridge, pp.272-294
- Bergson, H., 1903, "Einfuhrung in die Metaphysik", translated as "Introduction to Metaphysics", T.E. Hulme (trans.), available from <http://www.reasoned.org/dir/lit/int-meta.pdf>
- Blanchette, P., 2012, "The Frege-Hilbert Controversy", *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), E.N. Zalta (ed.), available from <http://plato.stanford.edu/archives/spr2014/entries/frege-hilbert/>
- Blumberg, A.E., & Feigl, H., 1974, “Introduction”, in M. Schlick, 1974, *General Theory of Knowledge*, A.E. Blumberg & H. Feigl (ed.), A.E. Blumberg (trans.), Open Court, LaSalle, IL
- Burge, T., 1979, “Individualism and the Mental”, *Midwest Studies in Philosophy* 4:1, pp.73-121
- Burgess, A., & Plunkett, D., 2013, “Conceptual Ethics”, *Philosophy Compass* 8:12, pp.1091-1101
- Cappelen, H., 2018 (forthcoming), “Conceptual Engineering: The Master Argument”, in A. Burgess, H. Cappelen, & D. Plunkett (eds.), *Conceptual Ethics and Conceptual Engineering*, Oxford University Press, Oxford, available from <http://hermancappelen.net/docs/ConceptualEngineeringTheMasterArgument.pdf>
- Carnap, R., 1927, "Eigentliche und Uneigentliche Begriffe", translated as "Proper and Improper Concepts", anonymous (trans.), available from https://www.princeton.edu/~hhalvors/teaching/phi520_f2012/EUB.pdf

- Carnap, R., 1928, *Der Logische Aufbau Der Welt*, translated as R. Carnap, 1967, *The Logical Structure of the World*, trans. George, R.A., Open Court, Chicago IL
- Carnap, R, 1932a, "Die physikalische Sprache als Unversalsprache der Wissenschaft", *Erkenntnis* 2, pp.432-465, translated as R. Carnap, 1934a, *The Unity of Science,* M. Black (trans.), Kegen, Paul, Trench, Teubner & Co, London
- Carnap, R., 1932b, "Uber Protokollsatze", *Erkenntnis* Vol.3, pp.215-228, translated as "On Protocol Sentences", R. Creath & R. Nollan (trans.), 1987, *Nous* 21:4, pp. 457-470
- Carnap, R., 1934b, *Logische Syntax der Sprache*, translated as *Logical Syntax of Language*, Amethe Smeaton (Countess von Zepplin) (trans.), 1937, Routledge, Trench, Trubner & Co Ltd, reprinted 2001, Routledge, New York
- Carnap, R., 1950, "Empiricism, Semantics, and Ontology", *Revue Internationale de Philosophie* 4, pp. 20-40, reprinted in 1956, Supplement to *Meaning and Necessity: A Study in Semantics and Modal Logic*, enlarged edition, University of Chicago Press, Chicago
- Chalmers, D., 2012, *Constructing the World*, Oxford University Press, Oxford
- Davidson, D., 1967, "Truth and Meaning", *Synthese* 17, pp.304-323
- Davidson, D., 1981, "A Coherence Theory of Truth and Knowledge", reprinted in D. Davidson, 2001, *Subjective, Intersubjective, Objective: Philosophical Essays, Volume 3,* Oxford University Press, Oxford, pp.137-158
- Davidson, D., 1982, "Empirical Content", reprinted in D. Davidson, 2001, *Subjective, Intersubjective, Objective: Philosophical Essays, Volume 3,* Oxford University Press, Oxford, pp.159-176
- Davidson, D., 1987, "Knowing One's Own Mind", *Proceedings and Addresses of the American Philosophical Association (1987)*, pp.441-458 reprinted in D. Davidson, 2001, *Subjective, Intersubjective, Objective: Philosophical Essays, Volume 3*, Oxford University Press, Oxford, pp.15-38
- Davidson, D., 1991, "Epistemology Externalized", *Dialectica* 45, pp.191-202, reprinted in D. Davidson, 2001, *Subjective, Intersubjective, Objective: Philosophical Essays, Volume 3,* Oxford University Press, Oxford, pp.193-204
- Engler, F.O., & Neuber, M., 2006, "Einleitung", in M. Schlick, 2006, *Uber die Reflexion des Lichtes in einer inhomogenen Schict & Raum und Zeit in der gegenwartigen Physik*, F. Stadler & H.J. Wendel (eds.), SpringerWienNewYork, Vienna, pp.9-52
- Feigl, H., 1938, "Moritz Schlick", *Erkenntnis* 7, pp.393-419, translated as "Moritz Schlick, A Memoir", M. Schlick, 1979a, *Moritz Schlick, Philosophical Papers, Volume I, [1909-1922]*, H. Mulder, & B.F.B. van de Velde-Schlick (eds.), P. Heath (trans.), D. Reidel Publishing Company, London, pp.xv-xxxviii

- Frege, G., 1980, *Philosophical and Mathematical Correspondence*, G. Gabriel, H. Hermes, F. Kambartel, C. Thiel, & A. Veraart (eds.), H. Kaal (trans.), Basil Blackwood Publisher, Oxford
- Friedman, M., 1987, "Carnap's Aufbau reconsidered", *Nous* 21:4, pp.521-545
- Geymonat, L., 1985, "Development and Continuity in Schlick's Thought", *Synthese* 64, pp.273-282
- Grice H.P., & Strawson, P.F., "In Defense of Dogma", *Philosophical Review* 65, pp.141-158
- Goldfarb, W., 1996, "The Philosophy of Mathematics in Early Positivism", in R.N. Giere & A.W. Richardson (eds.), *Origins of Logical Empiricism, Volume XVI*, University of Minnesota Press, Minneapolis, pp.213-230
- Hacker, P.M.S., 1986, *Insight and Illusion: Themes in the Philosophy of Wittgenstein*, revised edition, Oxford University Press, New York
- Haslanger, S., 2000, "Gender and Race: (What) Are They? (What) Do We Want Them To Be?", *Nous* 34:1, pp.31-55
- Helme, M., 1979, "An Elucidation of Tractatus 3.263", *The Southern Journal of Philosophy* 17:3, pp.323-334
- Hempel, H., 1935, "On the Logical Positivists' Theory of Truth", *Analysis* 2:4, pp.49-59
- Hilbert, D., 1899, *Grundlagen der Geometrie*, translated as D. Hilbert, 1950, *The Foundations of Geometry*, E.J. Townsend (trans.), Open Court Publishing, La Salle, IL
- Howard, D., 1993, "Einstein, Kant, and the Origins of Logical Empiricism", available from <https://www3.nd.edu/~dhoward1/Einstein,%20Kant%20and%20the%20Origins%20of%20Logical%20Empiricism.pdf>
- Hume, D., 1777, *Enquiries Concerning Human Understanding*, reprinted in D. Hume, 1975, *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, P.H. Nidditch (ed.), Oxford University Press, Oxford
- Jackson, F., 1982, "Epiphenomenal Qualia", *Philosophical Quarterly* 32:127, pp.127-136
- Keyt, D., 1964, "Wittgenstein's Picture Theory of Language", *The Philosophical Review* 73:4, pp.493-511
- Kripke, S., 1980, *Naming and Necessity*, Blackwell Publishing, Oxford
- Manninen, J., undated, "An Analysis of the 'Diktat für Schlick', Attributed to Wittgenstein", available from https://www.academia.edu/8266987/An_Analysis_of_the_Diktat_f%C3%BCr_Schlick_Attributed_to_Wittgenstein, accessed 15/03/19
- Monk, R., 1991, *Wittgenstein; The Duty of Genius*, Jonathan Cape Ltd, London

- Neurath, O., 1930, "Ways of the Scientific World-Conception", in O Neurath, 1983, *Philosophical Papers, 1913-1946*, D. Reidel Publishing Company, Boston, pp.32-47
- Neurath, O., 1931, "Physicalism", in O. Neurath, 1983, *Philosophical Papers, 1913-1946*, D. Reidel Publishing Company, Boston, pp.52-57
- Neurath, O., 1932, "Protokollsatze", *Erkenntnis* 3, pp.204-214, translated as "Protocol Sentences", in A.J. Ayer (ed.), 1959, *Logical Positivism*, G. Schick (trans.), The Free Press inc., New York, pp.199-208
- Neurath, O., 1934, "Radical Physicalism and the 'Real World'", in O. Neurath, 1983, *Philosophical Papers, 1913-1946*, D. Reidel Publishing Company, Boston, pp.100-114
- Newman, M.H.A, 1928, "Mr Russell's Causal Theory of Perception", *Mind* 37, pp.137-148
- Oberdan, T., 1996, "Postscript to Protocols", in R.N. Giere & A.W. Richardson (eds.), *Origins of Logical Empiricism, Volume XVI*, University of Minnesota Press, Minneapolis, pp.269-291
- Oberdan, T., 2013, "Moritz Schlick", *The Stanford Encyclopedia of Philosophy* (Summer 2013 Edition), E.N. Zalta (ed.), available from <http://plato.stanford.edu/archives/sum2013/entries/schlick>
- Pincock, C., 2005, "A Reserved Reading of Carnap's *Aufbau*", *Pacific Philosophical Quarterly* 86, pp.518-543
- Proops, I., 2013, "Wittgenstein's Logical Atomism", *The Stanford Encyclopedia of Philosophy* (Summer 2013 Edition), E.N. Zalta (ed.), available from <http://plato.stanford.edu/archives/sum2013/entries/wittgenstein-atomism>
- Putnam, H., 1973, "Meaning and Reference", *Journal of Philosophy* 70, pp. 699-711, reprinted in A.P. Martinich (ed.), 2008, *The Philosophy of Language*, 5th edition, Oxford University Press, Oxford, pp.306-313
- Putnam, H., 1960, "Do True Assertions Correspond to Reality?" in H. Putnam (ed.), 1975, *Philosophical Papers: Volume 2*, Cambridge University Press, Cambridge, pp. 70-84
- Quine, W.V.O., 1951, "Two Dogma's of Empiricism", *Philosophical Review* 60, reprinted in A.P. Martinich (ed.), 2008, *The Philosophy of Language*, 5th edition, Oxford University Press, Oxford, pp.63-76
- Ramsey, F.P., 1929, [1978], "Theories", *Foundations; Essays in Philosophy, Logic, Mathematics, and Economics*, Routledge & Kegan Paul, London, pp.101-125
- Ramsey, F.P., 1931, "Philosophy", reprinted in A.J. Ayer (ed.), 1959, *Logical Positivism*, The Free Press, New York, pp.321-326,
- Richardson, A.W., 1998, *Carnap's Construction of the World: The Aufbau and the Emergence of Logical Empiricism*, Cambridge University Press, Cambridge

- Ricketts, T., 1996, "Pictures, Logic, and the Limits of Sense in Wittgenstein's *Tractatus*", *The Cambridge Companion to Wittgenstein*, H. Sluga & D.G. Stern (eds.), Cambridge University Press, Cambridge, pp.59-99
- Rhees, R., 1969, "Note in Editing", in L. Wittgenstein, 1976, *Philosophical Grammar*, R. Rhees (ed.), A. Kenny (trans.), Basil Blackwell, Oxford, pp.487-490
- Russell, B., 1917, "The Relation of Sense-data to Physics", *Mysticism and Logic*, George Allen and Unwin Ltd., London, reprinted 1951, Barnes and Noble Books, Totowa, NJ, pp.108-131
- Schlick, M., 1908, *Lebensweisheit. Versuch einer Gluckseligkeitslehre*, München, cited in Blumberg & Feigl, 1974
- Schlick, M., 1909, "Das Grundproblem der Ästhetik in entwicklungsgeschichtlicher Beleuchtung", *Archiv für die gesamte Psychologie* 14, pp.102-132, translated as "The Fundamental Problem of Aesthetics Seen in an Evolutionary Light", M. Schlick, 1979a, *Moritz Schlick, Philosophical Papers, Volume I, [1909-1922]*, H. Mulder, & B.F.B. van de Velde-Schlick (eds.), P. Heath (trans.), D. Reidel Publishing Company, London, pp.1-24
- Schlick, M., 1910, "Das Wesen der Wahrheit nach der modernen Logik", *Vierteljahrsschrift fur wissenschaftliche Philosophie und Soziologie* 34, pp.386-477, translated as "The Nature of Truth in Modern Logic", M. Schlick, 1979a, *Moritz Schlick, Philosophical Papers, Volume I, [1909-1922]*, H. Mulder, & B.F.B. van de Velde-Schlick (eds.), P. Heath (trans.), D. Reidel Publishing Company, London, pp.41-103
- Schlick, M., 1912, *Uber die Reflexion des Lichtes in einer inhomogenen Schict*, reprinted in M. Schlick, 2006, *Uber die Reflexion des Lichtes in einer inhomogenen Schict & Raum und Zeit in der gegenwartigen Physik*, F. Stadler & H.J. Wendel (eds.), SpringerWienNewYork, Vienna
- Schlick, M., 1915, "Die philosophische Bedeutung des Relativitätspinzips", *Zeitschrift für Philosophie und philosophische Kritik* 159, pp.129-175, translated as "The Philosophical Significance of the Principle of Relativity", M. Schlick, 1979a, *Moritz Schlick, Philosophical Papers, Volume I, [1909-1922]*, H. Mulder, & B.F.B. van de Velde-Schlick (eds.), P. Heath (trans.), D. Reidel Publishing Company, London, pp.153-189
- Schlick, M., 1918, *Allgemeine Erkenntnislehre*, 1st edition, Springer, Berlin
- Schlick, M., 1920, *Space and Time in Contemporary Physics; Introduction to the Theory of Relativity and Gravitation*, 3rd edition, H.L. Brose (trans.), Clarendon Press, Oxford
- Schlick, M., 1925, *Allgemeine Erkenntnislehre*, 2nd edition, translated as M. Schlick, 1974, *General Theory of Knowledge*, A.E. Blumberg & H. Feigl (ed.), A.E. Blumberg (trans.), Open Court, LaSalle, IL

- Schlick, M., 1926, "Erleben, Erkennen, Metaphysik", *Kant Studien* 31, pp.146-158 translated as "Experience, Cognition, and Metaphysics", M. Schlick, 1979b, *Moritz Schlick, Philosophical Papers, Volume II, [1925-1936]*, H. Mulder, & B.F.B. van de Velde-Schlick (eds.), P. Heath (trans.), D. Reidel Publishing Company, London, pp.99-111
- Schlick, M., 1927, "Vom Sinn des Lebens", *Symposion* 1, pp.331-354, translated as "On the Meaning of Life", M. Schlick, 1979b, *Moritz Schlick, Philosophical Papers, Volume II, [1925-1936]*, H. Mulder, & B.F.B. van de Velde-Schlick (eds.), P. Heath (trans.), D. Reidel Publishing Company, London, pp.112-129
- Schlick, M., 1930a, "Die Wende der Philosophie", *Erkenntnis* 1, pp.4-11, translated as "The Turning-Point in Philosophy", M. Schlick, 1979b, *Moritz Schlick, Philosophical Papers, Volume II, [1925-1936]*, H. Mulder, & B.F.B. van de Velde-Schlick (eds.), P. Heath (trans.), D. Reidel Publishing Company, London, pp.154-160
- Schlick, M., 1930b, "Gibt es ein materiales Apriori?", translated as "Is There a Factual A Priori?", M. Schlick, 1979b, *Moritz Schlick, Philosophical Papers, Volume II, [1925-1936]*, H. Mulder, & B.F.B. van de Velde-Schlick (eds.), W. Sellars (trans.), D. Reidel Publishing Company, London, pp.161-170
- Schlick, M., 1930c, *Fragen der Ethik*, translated as M. Schlick, 1939, *Problems of Ethics*, D. Rynin (trans.) Prentice-Hall Inc., New York
- Schlick, M., 1932a,"Positivismus und Realismus", *Erkenntnis* 3, pp.1-31 translated as "Positivism and Realism", M. Schlick, 1979b, *Moritz Schlick, Philosophical Papers, Volume II, [1925-1936]*, H. Mulder, & B.F.B. van de Velde-Schlick (eds.), P. Heath (trans.), D. Reidel Publishing Company, London, pp.259-284
- Schlick, M., 1932b, "Form and Content. An Introduction to Philosophical Thinking", reprinted in M. Schlick, 1979b, *Moritz Schlick, Philosophical Papers, Volume II, [1925-1936]*, H. Mulder, & B.F.B. van de Velde-Schlick (eds.), D. Reidel Publishing Company, London, pp.285-369
- Schlick, M., 1934, "Über das Fundament der Erkenntnis", *Erkenntnis* 4, pp.79-99 translated as "On the Foundation of Knowledge", M. Schlick, 1979b, *Moritz Schlick, Philosophical Papers, Volume II, [1925-1936]*, H. Mulder, & B.F.B. van de Velde-Schlick (eds.), P. Heath (trans.), D. Reidel Publishing Company, London, pp.370-387
- Schlick, M., 1935a, “Facts and Propositions”, *Analysis* 2, pp.65-70 reprinted in M. Schlick, 1979b, *Moritz Schlick, Philosophical Papers, Volume II, [1925-1936]*, H. Mulder, & B.F.B. van de Velde-Schlick (eds.), D. Reidel Publishing Company, London, pp.400-404
- Schlick, M., 1935b, "Sur les ‘Constatations’", *Sur le Fondement de la Connaissance*, *Actualités Scientifiques et Industrielles* 289, pp.405-413, translated as "Introduction and On 'Affirmations'", M. Schlick, 1979b, *Moritz Schlick, Philosophical Papers, Volume II, [1925-1936]*, H. Mulder, & B.F.B. van de

Velde-Schlick (eds.), trans. P. Heath (trans.), D. Reidel Publishing Company, London, pp.405-413
- Schlick, M., 1935c, “De la Relation entre les Notions Psychologiques et les Notions Physiques”, *Revue de Synthese* 10, pp.5-26, translated as "On the Relation Between Psychological and Physical Concepts", M. Schlick, 1979b, *Moritz Schlick, Philosophical Papers, Volume II, [1925-1936]*, H. Mulder, & B.F.B. van de Velde-Schlick (eds.), W. Sellars (trans.), D. Reidel Publishing Company, London, pp.420-436
- Schlick, M., 1935d, “Sind die Naturgesetze Konventionen?”, *Induction et Probabilité*, Actes du Congres International de Philosophie Scientifique, Paris, translated as “Are Natural Laws Conventions?”, M. Schlick, 1979b, *Moritz Schlick, Philosophical Papers, Volume II, [1925-1936]*, H. Mulder, & B.F.B. van de Velde-Schlick (eds.), H. Feigl & M. Brodbeck (trans.), D. Reidel Publishing Company, London, pp.437-445
- Schlick, M., 1936, "Meaning and Verification", *The Philosophical Review*, No.45, pp.339-369, reprinted in M. Schlick, 1979b, *Moritz Schlick, Philosophical Papers, Volume II, [1925-1936]*, H. Mulder, & B.F.B. van de Velde-Schlick (eds.), D. Reidel Publishing Company, London, pp.456-481
- Schlick, M., 1986, *Die Probleme der Philosophie in ihrem Zusammenhang*, H.L. Mulder, A. Kox & R. Hegselmann (eds.), Suhrkamp, Frankfurt, translated as M. Schlick, 1987, *The Problems of Philosophy in their Interconnection*, Kluwer, Dordrecht cited in Uebel, 2007
- Stenius, E., 1964, *Wittgenstein's Tractatus; A Critical Exposition of its Main Lines of Thought*, Basil Blackwell and Mott Ltd., Oxford
- Uebel, T., 1992, *Overcoming Logical Positivism from Within: The Emergence of Neurath's Naturalism in the Vienna Circle's Protocol Sentence Debate*, Rodopi, Amsterdam
- Uebel, T., 2007, *Empiricism at the Crossroads: The Vienna Circle's Protocol-Sentence Debate Revisited*, Open Court, Chicago
- Waismann, F., 1938, “Vorwort”, in M. Schlick, 1938, *Gesammelte Aufsätze 1926-1936*, pp.vii-xxxi, translated as “Forward”, M. Schlick, 1979b, *Moritz Schlick, Philosophical Papers, Volume II, [1925-1936]*, H. Mulder, & B.F.B. van de Velde-Schlick (eds.), P. Heath (trans.), D. Reidel Publishing Company, London, pp.xiii-xxxiii
- Wellman, C., 1962, “Wittgenstein’s Conception of a Criterion”, *The Philosophical Review* 71:4, pp.433-447
- White, R.M., 2006, *Wittgenstein's Tractatus Logico-Philosophicus*, Continuum, London
- Wilson, G., & Shpall, S., 2016, "Action", *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), E.N. Zalta (ed.), available from <https://plato.stanford.edu/archives/win2016/entries/action/>

- Wittgenstein, L., 1922, *Tractatus Logico-Philosophicus*, C.K. Ogden (trans.), Kegan Paul, Trench, Trubner & Co., Ltd, London
- Wittgenstein, L., 1929, "Some Remarks on Logical Form", *Proceedings of the Aristotelian Society, Supplementary Volumes* 9, pp.162-171
- Wittgenstein, L., 1930, *Philosophical Remarks*, reprinted in R. Rhees (ed.), 1975, R. Hargreaves & R. White (trans.), Basil Blackwell, Oxford
- Wittgenstein, L., & Waismann, F., 2003, *The Voices of Wittgenstein; The Vienna Circle*, G. Baker (ed.), G. Baker, M. Mackert, J. Connolly & V. Politis (trans.), Routledge, London
- Wittgenstein, L., 1953, *Philosophical Investigations*, reprinted 2009, P.M.S. Hacker & J. Schulte (eds.), G.E.M. Anscombe, P.M.S. Hacker & J. Schulte (trans.), Blackwell Publishing Ltd, Oxford
- Wittgenstein, L., 1974, *Philosophical Grammar*, R. Rhees (ed.), A. Kenny (trans.), Basil Blackwell, Oxford
- Wright, C., 1984, “Second Thoughts About Criteria”, *Synthese* 58, D. Reidel Publishing Company, pp.383-405